EXACT NONPARAMETRIC TESTS FOR RANDOMIZED BLOCKS

By John E. Walsh[1]

Lockheed Aircraft Corporation

1. Summary. A class of nonparametric procedures for testing the statistical identity of treatments in randomized block experiments is suggested and discussed. The suggested procedures are squarely based on experimental within-block randomizations, and they may be chosen so as to have special power against particular alternatives. The blocks are assumed to be statistically independent but no assumption is made concerning dependence within the various blocks. The basic idea is to obtain from each block a statistic that is, under the null hypothesis, symmetrically distributed about zero and then to apply to the set of these statistics a nonparametric test of symmetry about zero. The observational data can be of any quantitative type.

2. Introduction. This paper considers experimental designs that are laid out in statistically independent blocks. If care is exercised, the blocks can usually be separated enough in distance, time, etc., to warrant the assumption of statistical independence.

Within a block, the assignment of the treatments investigated in that block can be of either a balanced or an unbalanced nature. For a given design, some blocks might be balanced and others unbalanced. The within-block assignments of treatments to locations are determined by a set of independent randomization processes as follows: the treatments of each block are partitioned into disjoint classes, to each class there is assigned a set of eligible locations within the block, and the assignments of treatments within a class to their eligible locations, for some classes, those of type A, are strictly random (all assignments equally likely), possibly dependent from class to class but independent from block to block. A block always contains at least one class of type A and each of these contains at least two treatments. For the remaining classes, those of type B, assignment to location may be random or fixed. The partitioning scheme, which may vary from block to block, is selected on the basis of the null hypothesis and the alternative hypotheses being investigated.

The most elementary type of situation considered is that where, for each block, the treatments (at least two per block) are not partitioned. Then all the classes (one per block) are of type A and the null hypothesis asserts that, for each block, the joint distribution of the observations is invariant under all permutations of the names of treatments within each block. Also, within a block, the locations are eligible for all the treatments and are randomly assigned to these treatments.

This elementary situation can be generalized in several respects through the

Received September 30, 1957; revised March 28, 1958.

[1] Now with the System Development Corporation, Santa Monica, California.

use of partitioning. First, various combinations of treatments can be selected for comparison within a block. Second, the eligible locations associated with the classes of the partition can be chosen in many ways. Finally, the part of the null hypothesis that pertains to a given block need not consider all the treatments of this block. In fact, only the treatments of the partition classes of type A are considered. The null hypothesis asserts that the joint distribution of the observations for a block is invariant under all permutations of the names of treatments within a partition class for all the classes of type A. That is, excluding partition classes of type B, under the null hypothesis the treatments within a partition class have identical joint probability properties.

The procedure of including treatments in a block which are not considered in the part of the null hypothesis pertaining to this block serves a useful purpose. For the situations of this paper, a treatment is included in the experiment for one or both of two reasons. First, the question of whether this treatment is identical with a specified one or more other treatments can be of interest. This type of relation is considered in selection of the null hypothesis H_0. Second, there can be interest in a given form of interrelation that might exist between this treatment and specified other treatments when the null hypothesis is false. This second type of relation is considered in identifying the alternative hypotheses that are of principal interest. For each block, those treatments which are included exclusively for the second reason are placed in one or more partition classes of type B and are not considered in the part of the null hypothesis that is associated with this block. The reason for using more than one partition class for these treatments (e.g., a separate partition class for each treatment) is that they may not have the same set of eligible locations.

The choice of the eligible locations for the various classes of the partition is at the discretion of the experimenter. Often this freedom of choice in specifying eligible locations can be exploited to obtain a more efficient experiment. As an example, for the treatments of some partition classes, location in one part of the block may be more important than location in other parts, because of a special condition that exists in this part. There is great freedom in selecting eligible locations for treatments, subject to the condition that the size of each set of eligible locations is at least as great as the number of treatments that could be assigned to this set. In particular, two or more classes of the partition might have over-lapping sets of eligible locations. For this case, it is convenient, but not necessary, to require that if any two classes of the partition are to have at least one eligible location in common they must have the same set of eligible locations. Then, when two or more of the partition classes have the same set of eligible locations, their treatments can be handled as a group in performing the random assignment to eligible locations. This "grouping and then assigning" procedure greatly simplifies the random assignment scheme for situations of this nature. If desired, a specified location can be assigned to each treatment of a partition class of type B.

To perform the test, a statistic is specified for each block. This statistic depends on all the treatments for this block but not on those for any of the other blocks. These statistics are chosen so that they have symmetrical distributions about zero when the null hypothesis is true. They are also chosen so that the test is sensitive to the alternative hypotheses that are emphasized. The forms of the statistics can vary from block to block and this freedom can sometimes be exploited by tailoring the statistics to the special situations that exist for the blocks. Because of the broad nature of the situations considered, no generally applicable rules can be stated for choosing the block statistics so as to emphasize the alternative hypotheses of major interest. However, in many cases a reasonable selection can be made on an intuitive basis. Some examples of the selection of block statistics are given in Section 4. Since the observations are independent between blocks, the block statistics are independent and also, under the null hypothesis, have symmetrical distributions about zero. Consequently the null hypothesis can be tested by use of an appropriate nonparametric test of symmetry about zero. References to some nonparametric tests of symmetry about zero are given in Section 3.

No quantitative attempt is made to evaluate the efficiencies of the tests that can be obtained on the basis of this paper. The great freedom allowed in selecting the treatment partition classes, the eligible locations, and the block statistics, combined with the myriad of possible alternative hypotheses and the different kinds of tests that could be used, make such an investigation infeasible. Qualitative considerations, however, hint that in many cases the efficiency should be reasonably high if the eligible assignment locations and the block statistics are chosen so that the alternative hypotheses of major interest are emphasized. For example, if the normality model for experimental design holds and the treatment comparisons are linear, the best test is that based on the appropriate t-statistic. A situation of this nature was examined in [1], under conditions that represent a special case of the results of this paper. The tests based on the block statistics used in [1] were found to have efficiencies that are only in the neighborhood of 60–70 percent for the case of normality and linear comparisons. However, if the most appropriate treatment comparison for the alternative hypotheses of interest is not linear, suitable selection of the block statistics so as to emphasize these alternative hypotheses may furnish the basis for nonparametric tests that are much more efficient than the best t-tests based on linear comparisons. Of course, if the block statistics and the eligible locations are poorly chosen, a test of this type can have a very low efficiency.

It is no loss of generality to suppose, for purposes of formal theory, that the treatments of a block are possibly different, or at least have different names; also to suppose that different blocks can contain different treatments. Situations where treatments are replicated or where the same treatments occur in several blocks represent special cases of this general situation.

The null hypothesis of treatment equivalence for specified treatment partitions can be generalized. Instead of specifying that, for the partition classes whose treatments are named in H_0 (i.e., class A), the treatments of a partition class have identical probability properties, the null hypothesis could assert that this is the case if specified transformations are made of these values. By use of transformations, the classes of null hypotheses that are available for consideration and of alternative hypotheses that are emphasized by elementary form block statistics can be greatly extended.

In Section 3 the permissible forms for the block statistics, verification of their properties under the null hypothesis, and a statement of how the test makes use of these statistics are presented. Section 4 is titled Block Statistic Selection. Two examples are given to illustrate the intuitive selection of block statistics so as to emphasize specified alternative hypotheses.

3. Results. The principal purpose of this section is to show that a properly chosen block statistic is symmetrically distributed about zero under H_0. Consequently, all the notation occurring in the derivation applies to an arbitrary but specified block. The random method used to assign the treatments of a partition to their eligible locations is described in Section 2 and is not repeated here.

Suppose that treatments $1, 2, \cdots, I$ occur in the block, and that these are partitioned into $T + 1$ classes. The first T of these classes are of type A and the last of type B. (If there are several partition classes of type B, nothing is lost by throwing them together and working conditionally on whatever random assignments to location may have been made for such type B classes.) The tth set contains $k(t) - k(t-1)$ treatments, with $k(0) = 0$ and $k(T+1) = I$. The treatments for the tth set are denoted by

$$i_{k(t-1)+1}, i_{k(t-1)+2}, \cdots, i_{k(t)} \qquad (t = 1, \cdots, T+1).$$

The partitioning is done so that sets $1, \cdots, T$ are the partition classes which are used in the part of the null hypothesis pertaining to this block (i.e., class A), while the remaining set contains all the treatments that are in partition classes which do not appear in H_0 (class B). In terms of this notation, the null hypothesis associated with this block asserts that all the treatments of the tth set have identical joint probability properties with respect to the experiment for $t = 1, \cdots T$.

Let the random variable $y(i)$ represent the observable result for the ith treatment ($i = 1, \cdots, I$), where the probability effects from the randomization and the experimentation are combined to obtain the joint distribution of $y(1), \cdots, y(I)$. For $1 \leqq t \leqq T$, let ϕ_t denote an arbitrary permutation of the numbers $i_{(k-1)+1}, \cdots, i_{k(t)}$; also let ϕ_{T+1} be the identity transformation for $i_{k(T)+1}, \cdots, i_I$. Use

$$F\{y[i_{k(t-1)+1}], \cdots, y[i_{k(t)}]; 1 \leqq t \leqq T+1\}$$

to denote the joint cumulative distribution function (cdf) for $y(1), \cdots, y(I)$. Then, on the basis of the randomization scheme and the null hypothesis,

$$\begin{aligned} &F\{y[i_{k(t-1)+1}], \cdots, y[i_{k(t)}]; 1 \leqq t \leqq T+1\} \\ &\qquad \equiv F\{y[\phi_t(i_{k(t-1)+1})], \cdots, y[\phi_t(i_{k(t)})]; 1 \leqq t \leqq T+1\}. \end{aligned} \tag{1}$$

That is, under H_0, the joint cdf of $y(1), \cdots, y(I)$ is invariant under all possible permutations within each of the T sets of treatments that are considered in the null hypothesis. No moments of any order are assumed to exist for the $y(i)$.

A block statistic is a function of $y(1), \cdots, y(I)$ which is denoted by

$$g\{y[i_{k(t-1)+1}], \cdots, y[i_{k(t)}]; 1 \leqq t \leqq T+1\}.$$

This function, which is chosen so as to not be identically zero for all values of the $y(i)$, is required to have the property that there exists a set of permutations $\phi_1, \cdots, \phi_T, \phi_{T+1}$, where ϕ_{T+1} is the identity permutation, such that

$$g\{y[i_{k(t-1)+1}], \cdots, y[i_{k(t)}]; 1 \leqq t \leqq T+1\}$$
$$\equiv -g\{y[\phi_t(i_{k(t-1)+1})], \cdots, y[\phi_t(i_{k(t)}]; 1 \leqq t \leqq T+1\}.$$

But, on the basis of relation (1),

$$g\{y[i_{k(t-1)+1}], \cdots, y[i_{k(t)}]; 1 \leqq t \leqq T+1\}$$

and

$$g\{y[\phi_t(i_{k(t-1)+1})], \cdots, y[\phi_t(i_{k(t)})]; 1 \leqq t \leqq T+1\}$$

have the same distribution. Thus $-g$ has the same distribution as g if the null hypothesis holds; consequently, under H_0, the block statistic g has a probability distribution that is symmetrical about zero.

Since, by hypothesis, the observations are statistically independent between blocks, the block statistics are a set of independent random variables with distributions that are symmetrical about zero if the null hypothesis is true. A wide variety of nonparametric procedures are available for testing the symmetry of populations about zero. These include the signed-rank test of Wilcoxon [2], [3], [4], [5], the Fisher test [6], Nair's test [7], a comprehensive set of tests by Hemelrijk [8] and by van Eeden and Benard [9], and the results of [10], [11]. If the distributions of the block statistics are not all continuous, tests based on the assumption of continuity can be validly used by appropriate randomization of ties. Alternately, some of the tests are valid for both discrete and continuous populations (see, e.g., [6], [7], [8], [9]).

The efficiency of this testing procedure depends on the test used, the forms of the block statistics, the partitioning scheme, and the choice of eligible locations for treatments. In particular, the forms of the block statistics have a strong influence on which alternative hypotheses are emphasized. The next section considers intuitively the problem of choosing the forms of the block statistics so as to emphasize specified types of alternative hypotheses.

4. Block statistic selection. The great freedom in selecting the forms for the block statistics allows so many types of situations to arise that no general rule for the selection of these statistics seems to be available. The alternative hypotheses which are eligible for consideration are of such a wide class that determination of a general method of selecting a block statistic so as to emphasize an arbitrary but specified alternative hypothesis (hypotheses) does not appear to be feasible. However, a reasonable (but not necessarily preferable) selection can often be made on the basis of judgment combined with intuitive considerations. Two examples are given which illustrate the intuitive method of selecting block statistics and which are somewhat typical of situations of practical interest. In these examples, the same form is considered to be usable for all blocks. However, since the considerations are on the basis of a single block, these considerations also apply to cases where the forms may change from block to block.

First example: Let $I = 8$ and suppose that the null hypothesis asserts that treatment 1 is equivalent to treatment 2 and that treatments 3–6 are equivalent. Treatments 7 and 8 do not occur in the statement of the null hypothesis. The three alternative hypotheses of principal interest are

H_1 : The value of treatment 1 tends to be larger than that of treatment 2, but small deviations are not important.

H_2 : The average of the values of treatments 3 and 4 tends to be smaller than the average of the values of treatments 5 and 6.

H_3 : The value for treatment 1 minus that for treatment 2 tends to be negative and simultaneously the average of the values of treatments 3–8 tends to exceed 10.

If all of H_1–H_3 hold, or if at least one holds in a strong fashion and neither of the one-sided H's holds strongly in a negative sense, it is highly desirable that the null hypothesis be rejected.

For this case, use of the function

$$\{y[1] - y[2]\}^3 - \tfrac{1}{2}\{y[3] + y[4] - y[5] - y[6]\} \\ - \tfrac{1}{6}\{y[3] + \cdots + y[8] - 60\}\, sgn\,\{y[1] - y[2]\}$$

for g, combined with an appropriate one-sided test for symmetry about zero (which is sensitive to large positive values of the variable) might be satisfactory. The first term accounts for the alternative H_1, the second term for H_2, and the third term for H_3. The permutations

$$\phi_1 : 1 \leftrightarrow 2 \qquad \phi_2 : 3 \leftrightarrow 5 \quad \text{and} \quad 4 \leftrightarrow 6$$

result in a change of sign for g.

Second example. Let $I = 4$. The null hypothesis asserts that all four treatments are equivalent. The alternative hypothesis of principal interest is

H_1 : The sign of the value of treatment 1 minus that of treatment 2 tends to be the same as that of the value of treatment 3 minus that of treatment 4. Also the magnitude of the difference involving treatments 1 and 2 tends to exceed that of the difference involving treatments 3 and 4.

For this case, use of the function

$$\{y[1] - y[2]\}/\{y[3] - y[4]\}$$

for g, combined with an appropriate one-sided test of symmetry about zero (which is sensitive to large positive values of the variables), would appear to be suitable. The permutation

$$\phi_1 : 1 \leftrightarrow 2,\ 3 \leftrightarrow 3, \quad \text{and} \quad 4 \leftrightarrow 4$$

results in a change of sign for g.

REFERENCES

[1] JOHN E. WALSH, "Some nonparametric tests for Student's hypothesis in experimental designs," *Jour. Amer. Stat. Assoc.*, Vol. 47 (1952), pp. 401–415.

[2] FRANK WILCOXON, "Individual comparisons by ranking methods," *Biom. Bull.*, Vol. 1 (1945), pp. 80–83.

[3] FRANK WILCOXON, "Individual comparisons of grouped data by ranking methods," *Jour. Econ. Entomology*, Vol. 39 (1946), pp. 269–270.

[4] FRANK WILCOXON, "Probability tables for individual comparisons by ranking methods," *Biometrics*, Vol. 3 (1947), pp. 119–122.

[5] FRANK WILCOXON, *Some Rapid Approximate Statistical Procedures*, American Cyanamid Co., Stamford Research Laboratories, 1949.

[6] R. A. FISHER, *The Design of Experiments* (5th Ed.), Hafner Publishing Co., Inc., N. Y., 1949, pp. 43–47.

[7] K. R. NAIR, "The median in tests by randomization," *Sankhyā*, Vol. 4 (1940), pp. 543–550.

[8] J. HEMELRIJK, "A family of parameterfree tests for symmetry with respect to a given point I, II," *Proc. Kon. Ned. Akad. v. Wet., Ser. A*, Vol. 53 (1950), pp. 945-955 and pp. 1186–1198; also *Indag. Math.*, Vol. 12 (1950), pp. 340–350 and 419–431.

[9] CONSTANCE VAN EEDEN AND A. BENARD, "A general class of distribution-free tests. for symmetry containing the tests of Wilcoxon and Fisher I, II, III," *Proc. Kon. Ned. Akad. v. Wet., Ser. A*, Vol. 60 (1957), pp. 381–408; also *Indag. Math.*, Vol. 19 (1957).

[10] JOHN E. WALSH, "Some significance tests for the median which are valid under very general conditions," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 64–81.

[11] JOHN E. WALSH, "Applications of some significance tests for the median which are valid under very general conditions," *Jour. Amer. Stat. Assoc.*, Vol. 44 (1949), pp. 342–355.