

Permutation Techniques Based on Euclidean Analysis Spaces: A New and Powerful Statistical Method for Ecological Research

Mario E. Biondini^a, Paul W. Mielke, Jr.^b and Edward F. Redente^c

^a*Department of Animal and Range Science, North Dakota State University, Fargo, North Dakota 58105;*

^b*Department of Statistics, Colorado State University, Fort Collins, Colorado 80523; and*

^c*Department of Range Science, Colorado State University, Fort Collins, Colorado 80523, USA*

Abstract

The statistical procedures that are most widely used in ecological population and community research belong to the family of parametric methods. Embedded in these procedures are assumptions about the normal distribution of the underlying population, homogeneity of variances and linear response patterns. One of the problems encountered in ecological and vegetation studies, however, is that these assumptions are very difficult, if not impossible, to meet. In addition, a very serious shortcoming of the most widely used statistical methods is the lack of congruence between the geometry of the data space, which is for the most part Euclidean, and the analysis space, which in the standard parametric tests and most of the nonparametric tests, is not Euclidean. In ecological and vegetation studies, the combination of a failure to meet model assumptions and a lack of congruence between the geometries of the data space and the analysis space can lead, as shown in this paper, to gross errors in data interpretation and hypothesis testing.

A new and powerful statistical technique (MRPP) is presented in this paper which is free from assumptions about the underlying distribution model of the population under analysis, can easily handle nonlinear data structures and more importantly meets the congruence principle (a common geometry for both the data and analysis spaces). The theoretical formulations for MRPP and its randomized block design counterpart MRBP as well as their relationship to other statistical methods are outlined in the first part of a paper. This is followed by computer algorithms and programs needed for their implementation as well as a series of detailed examples which demonstrate major advantages of MRPP and MRBP over the currently most widely used statistical methods. Computer programs are available from the authors free of charge (send a blank diskette).

1. Introduction

The strict application of the scientific method to ecology in general and vegetation science, in particular, is critical if both disciplines are to advance significantly our knowledge of natural and introduced ecosystems as well as remove the stigma of 'soft science' that has been often applied to ecology (Simberloff 1981, Peters 1980). The full application of the scientific method not only implies the inductive route, that is what we generally call data interpretation, but more critically the hypothesis-deductive route. From the inductive approach we are able to generate hypotheses whose function are to explain in a cohesive and comprehensive way why a particular pattern in our observations has occurred. The next step involves the development of critical predictions from our

general hypothesis that are to be experimentally tested (Fig. 1). In almost all the cases involving vegetation studies it is impractical and/or impossible to analyze the entire population. This situation has lead us to make inferences about ecosystem properties based upon a limited number of samples forcing us to rely heavily on statistical procedures. The statistical procedure therefore becomes a key component in the acceptance or rejection of a hypothesis.

The statistical procedures that are widely used in ecology belong to the family of parametric methods. These methods assume an underlying distribution model for the data which is characterized by a given set of parameters. The test then involves the calculation of the parameters from the sampled observations and an estimation of the probability of obtaining that particular set of

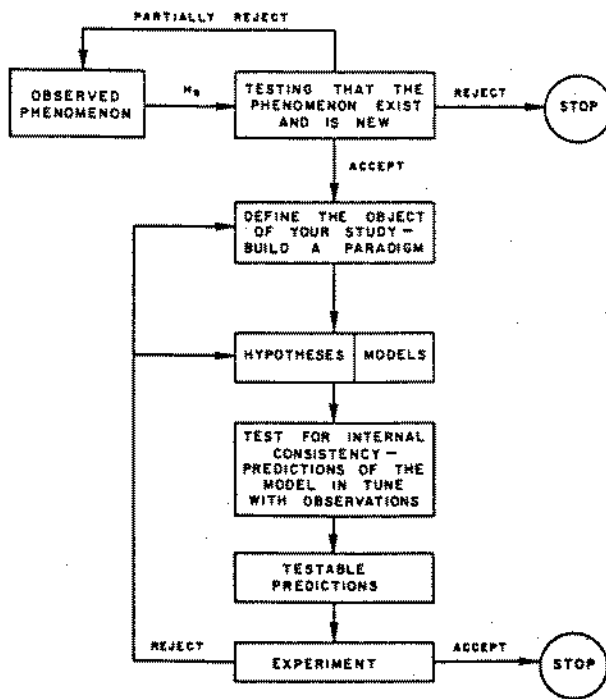


Fig. 1. A diagram showing the sequence of inductive and hypothetical-deductive steps on the scientific method. Note: H_0 = null hypothesis.

parameter values if all the samples had been taken from the same population (null hypothesis). The majority of the tests that are commonly available (e.g., t-test, Hotelling T^2 , analysis of variance (ANOVA), multivariate analysis of variance (MANOVA), multiple regression, canonical correlation, discriminant analysis) involve the estimation and testing of means and are characterized by the following assumptions:

- The underlying distribution model for a population is assumed to be normal or multivariate normal with equal (homogeneous) variance or an equal variance-covariance matrix. The normal distribution (and its counterpart the multivariate normal) are totally characterized by the mean and variance (mean vector and variance-covariance matrix). As a result of our assumption of equal variance (variance-covariance matrix), all the statistical procedures listed above (ANOVA, MANOVA, etc.) are mean based tests.
- A second and very important assumption made in all the above mentioned tests is that the underlying distribution models are linear.
- The methods used to calculate the pertinent statistics involve least squares. As a consequence, the inherent nature of the analysis space is non-metric.

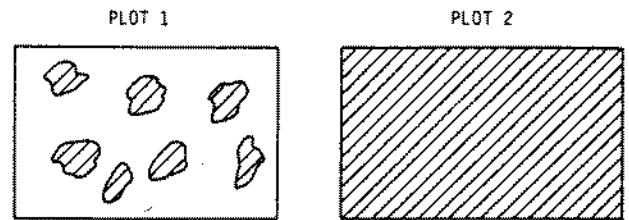


Fig. 2. Plot 1 and Plot 2 have equal biomass but different distribution, uniform in Plot 2 scatter in Plot 1.

In ecological field studies, all of these assumptions are very difficult if not impossible to meet. The underlying distribution model of the population, most of the time, is not known and in many cases is definitely not normal (for example, precipitation distribution in semi-arid regions is better approximated by an exponential distribution (Sala *et al.* 1982)). In the case of univariate analyses, this may be less of a problem because most univariate distributions can be brought closer to a normal distribution by an adequate transformation. However, this approach is difficult when multivariate distributions are involved. A most serious constraint, however, is the assumed equality of variances (variance-covariance matrix) of the underlying models. There are many cases of vegetation analyses when comparisons are made between vegetation communities that have different patterns of plant distribution (Fig. 2) or between particular vegetation treatments which may result in significant alterations in plant distribution patterns. Under these circumstances the assumption of equal variances (variance-covariance matrices) does not hold.

The linear assumptions involved in ANOVA, MANOVA or other standard statistical tests may be very difficult to meet. As mentioned by Orłóci (1979), linear relationships are rarely met in ecological systems and furthermore most of the ecologically interesting data is intrinsically tied up with non-linear correlations.

The least square solution method that is used in the standard statistical methods (ANOVA, MANOVA, etc.) may present another serious problem in ecological research, this time because of major differences between the geometry of the data space and the geometry of the analysis space. Let us define Δ_{ij} as the distance between points i and j in a given space. A space is called a metric space if (a) $\Delta_{ij} \geq 0$ and $\Delta_{ii} = 0$; (b) $\Delta_{ij} = \Delta_{ji}$ (symmetry); and (c) $\Delta_{ij} + \Delta_{jk} \geq \Delta_{ik}$ (triangle inequality). The most familiar metric space is the ordinary Euclidean space. All of the standard statistical methods belong to the family of generalized linear models where the estimation of parameters involve the minimization of functions such as

$$\sum_{i=1}^n (Y_i - X_i \Theta)^2$$

where Y_i and X_i are the dependent and independent observations and Θ is the parameter in question. The distance measure associated with this approach does not meet the triangle inequality and consequently the geometry of the analysis space is nonmetric (Mielke 1986). The data space of essentially all ecological studies (which may involve transformations) is usually perceived as an Euclidean space of measurements (e.g., biomass, height, cover and weight), the most common metric space. This difference between the geometry of the analysis space and the data space imply that the congruence principle (Mielke 1985, 1986) is not satisfied. This violation of the congruence principle is the root of commonly encountered contradictions between one's intuition (from observation) and inferential results based on well known statistical methods (such as ANOVA, MANOVA, etc.).

As was mentioned before, all of the traditional statistical methods are mean based methods. In many instances pertaining to ecology, the mean value may be less important than other parameters. In Fig. 2, for instance, both communities may have similar mean values for biomass, but this could be ecologically less important than how this biomass is distributed over the landscape. As a consequence, a statistical procedure that tests for biomass concentration rather than averages would be more meaningful. In a normal distribution the mean and the median are equal. If the distribution of the data is both non-normal and non-symmetric, however, the median value in some circumstances could be ecologically more important than the mean value.

The above discussion clearly indicates that ecological and vegetation research is in need of a strong statistical inference method that: (1) is free from assumptions about a normal (or multivariate normal) distribution, equality of variances (or variance-covariance matrices) and linear relationships; (2) satisfies the congruence principle (i.e., the analysis space is a Euclidean space); and (3) is flexible enough to allow for the testing of a variety of structural differences (means, medians, data concentration or dispersion, etc.) between the populations in question. In Section 2 we describe statistical methods, *multi-response permutation procedures* (MRPP) and *its randomized block design analog* (MRBP), that meet the criteria proposed above. In this section we also discuss the interconnection of MRPP and MRBP with ANOVA and MANOVA tests when nonmetric analysis spaces are involved. Section 3 gives an outline of algorithms and

computer programs developed to implement MRPP and MRBP. Several examples with real and artificial data are presented in Section 4 to highlight major characteristics of MRPP and MRBP and how they compare with results derived from traditional statistical methods. A summary and comments about the advantages of MRPP and MPBP for ecological research is presented in Section 5. For those not interested in the mathematical foundations of MRPP and MRBP or the corresponding algorithms, we suggest that the reader proceed to Section 4.

2. The new statistical procedures

2.1. Methodological description (MRPP)

Let $\Omega = [\omega_1, \dots, \omega_N]$ be a finite population of N objects, let $X_i = (x_{i1}, \dots, x_{ir})$ denote r commensurate response measurements (these might be functions of response measurements or residuals adjusted by predictors) for object $\omega_i (i = 1, \dots, N)$, and let S_1, \dots, S_{g+1} designate an exhaustive partitioning of the N objects comprising Ω into $g+1$ disjoint groups. The group S_{g+1} encompasses those objects that are not classified in the g groups, in other words, it represents an excess group. In most cases S_{g+1} will be empty but it can be very important in the testing of certain hypotheses as we will see in Section 4. Also let Δ_{ij} be a symmetric distance function value of the response measurements associated with the objects ω_i and ω_j . The statistic underlying MRPP is given by

$$\delta = \sum_{i=1}^g C_i \zeta_i \tag{2.1.1}$$

where

$$\zeta_1 = \left(\frac{n_i}{2}\right)^{-1} \sum_{I < J} \Delta_{ij} \Phi_i(\omega_I) \Phi_j(\omega_J) \tag{2.1.2}$$

is the average distance function value for all distinct pairs of objects in group $S_i (i = 1, \dots, g)$, $n_i > 2$ is the number of a *priori* classified objects in group $S_i (i = 1, \dots, g)$, $K = \sum_{i=1}^g n_i$, $n_{g+1} = N - K$ is the number of remaining (unclassified) objects in the excess group S_{g+1} , $\sum_{I < J} \Phi_I(\omega_I) \Phi_J(\omega_J)$ is the sum over all I and J such that $1 < I < J < N$, $\Phi_I(\omega_I)$ is 1 if ω_I belongs to S_i and 0 otherwise, $C_i > 0 (i = 1, \dots, g)$, and $\sum_{i=1}^g C_i = 1$. The C_i value generally used is $C_i = n_i/K$. Discussed later is the impact of using alternative C_i values. The underlying permutation distribution of δ (the null hypothesis) assigns equal probabilities to the

$$M = N! / \left(\prod_{i=1}^{g+1} n_i ! \right)$$

possible allocations of the N objects to the $g + 1$ disjoint

groups. The mean, variance and skewness of δ under the null hypothesis are denoted respectively by μ_δ , σ_δ^2 and γ_δ (Mielke 1984). Under the null hypothesis, preliminary findings (Mielke *et al.* 1976, Mielke 1978, 1979) indicated some situations when the asymptotic distribution of $N(\delta - \mu_\delta)$ is nondegenerate with the skewness of δ being substantially negative. Based on results due to Sen (1970, 1972), O'Reilly and Mielke (1980) presented general theorems for the multivariate case of MRPP which characterize situations in which the distribution of $N^{1/2}(\delta - \mu_\delta)$ is asymptotically normal under the null hypothesis. More recently, Brockwell *et al.* (1982) presented theorems for the univariate case of MRPP which delineate distributions for situations (probably the most important situations) in which the nondegenerate distribution of $N(\delta - \mu_\delta)$ is not asymptotically normal under the null hypothesis. The multivariate generalizations of the results by Brockwell *et al.* (1982) and special situations analogous to those considered by Mielke and Sen (1981) for linear rank statistics are open questions which require further attention.

The symmetric distance function (Δ_{IJ}) is extremely important since it defines the structure of the underlying analysis space of MRPP. The form of the symmetric distance functions considered in this paper will be confined to

$$\Delta_{IJ} = \left(\sum_{k=1}^r |x_{kI} - x_{kJ}|^p \right)^{1/p} \quad (2.1.3)$$

where $p > 1$ and $v > 0$ (p is not relevant when $r = 1$). In particular, the underlying analysis space of MRPP is nonmetric when $v > 1$ (i.e., the triangle inequality property of a metric space fails) and is metric when $v \leq 1$ (a distorted metric space when $v < 1$). The analysis space of MRPP is a Euclidean space when $p = 2$ and $v = 1$. While the validity of a permutation test is not affected by these geometric considerations, the rejection region of any test is highly dependent on the underlying geometry. Thus a geometry problem (i.e., either a nonmetric or a distorted metric space) will affect the power of a permutation test. The results of a permutation test will surely be misleading if the rejection region is incomprehensible. As a consequence, the choice of $p = 2$ and $v = 1$ is recommended for routine applications. This may be a controversial conclusion since, as subsequently demonstrated, the majority of permutation tests presently used in routine applications are based on $v = 2$.

Since the symmetric distance functions considered in this presentation are of the type described in the previous paragraph, small values of δ imply a concentration of response measurements within the g groups. Thus the P-value for a realized value of δ (say δ_0) is the probabil-

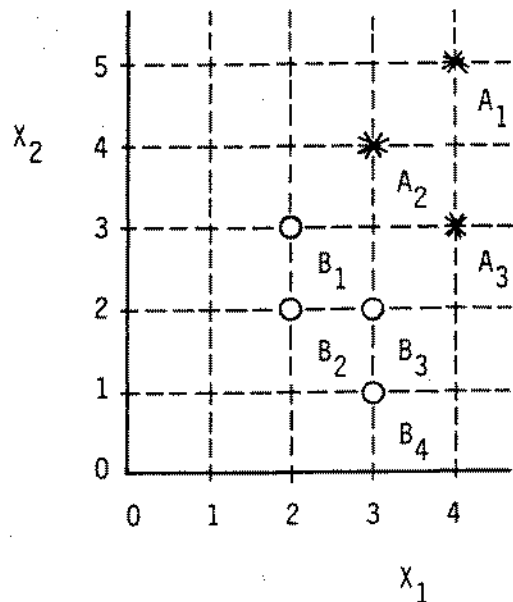


Fig. 3. Scatter diagram showing the points of the two subgroups (A and B) plotted as a function of the measured responses X_1 and X_2 .

ity statement given by $P(\delta < \delta_0)$.

As an example, let's look at a case where we have two groups of objects (A and B) from which two measured responses (x_1 and x_2) have been made (Fig. 3). The seven points of Fig. 3 imply there are $\binom{7}{2} = 21$ distinct pairs of points and consequently 21 distances must be computed. These 21 distances are listed in Table 1 and ordered from the lowest to highest value. The distances Δ_{IJ} are in this case calculated with the use of $p = 2$ and $v = 1$ in (2.1.3). Table 1 confirms the visual impression of clustering since the distance between points of a common group tend to be smaller than the distances between points of different groups. The values for ζ_A , ζ_B and M in this case are:

$$\zeta_A = (1/3) \sum_A \Delta_{I,J} = 1.6095$$

$$\zeta_B = (1/6) \sum_B \Delta_{I,J} = 1.3441$$

$$\delta = (3/7) \zeta_A + (4/7) \zeta_B = 1.4578$$

$$M = 7!/(3! 4!) = 35$$

Since $M = 35$ for this example, 35 values of δ can be obtained by enumerating all the possible 35 partitions. These 35 values of δ are listed in Table 2 and are ordered from the lowest to highest value. We see that the observed statistic ($\delta_0 = 1.4578$) obtained for the realized partition

Table 1. Ordered distances between all 21 pairs of the seven points shown in Fig. 4 where distances between points in either subgroup A or subgroup B are indicated by crosses or circles, respectively.

Rank	Points	Distance
1	B ₁ B ₂	1.000 (o)
2	B ₂ B ₃	1.000 (o)
3	B ₃ B ₄	1.000 (o)
4	A ₁ A ₂	1.414 (x)
5	A ₂ A ₃	1.414 (x)
6	A ₃ B ₁	1.414
7	A ₃ B ₃	1.414
8	B ₁ B ₃	1.414 (o)
9	B ₂ B ₄	1.414 (o)
10	A ₁ A ₃	2.000 (x)
11	A ₂ B ₃	2.000
12	A ₃ B ₁	2.000
13	A ₂ B ₂	2.236
14	A ₃ B ₂	2.236
15	A ₃ B ₄	2.236
16	B ₁ B ₄	2.236 (o)
17	A ₁ B ₁	2.828
18	A ₂ B ₄	3.000
19	A ₁ B ₃	3.162
20	A ₁ B ₂	3.606
21	A ₁ B ₄	4.123

(A and B) is indeed unusual since each of the remaining 34 values is greater than δ_0 . If all partitions could have occurred with equal chance (the null hypothesis), then the observed significance level or P-value is $1/35 = 0.0286$. Thus, we would accept the realized partition (A and B) as significant at the 2.8% level of significance.

Although an efficient algorithm for calculating the exact P-value for an observed value of δ has been developed (Berry and Mielke 1984), this procedure is extremely expensive when M is large (e.g., $M > 10^6$).

Table 2. Ordered values of δ for all 35 partitions of the seven points shown in Fig. 4 into groups (A and B) having fixed sized $n_A = 3$ and $n_B = 4$.

Rank	Value	Rank	Value
1	1.4578	19	2.1381
2	1.5421	20	2.1480
3	1.6939	21	2.1591
4	1.7505	22	2.1646
5	1.8389	23	2.1709
6	1.8547	24	2.1740
7	1.8935	25	2.1769
8	1.9898	26	2.1891
9	1.9915	27	2.1939
10	1.9988	28	2.2025
11	2.0060	29	2.2169
12	2.0157	30	2.2258
13	2.0176	31	2.2280
14	2.0522	32	2.2470
15	2.0575	33	2.2518
16	2.0829	34	2.2812
17	2.0944	35	2.2935
18	2.1158		

Since $n_1 = n_2 = 12$ and $N = 24$ yields $M = 2.7 \times 10^6$ or $n_1 = n_2 = n_3 = 6$ and $N = 18$ yields $M = 1.7 \times 10^7$, the need for an approximation procedure becomes essential for cases involving even relatively small sample sizes. Because the distribution of δ under the null hypothesis may differ substantially from a normal distribution for either small, moderate or extremely large sample sizes (Mielke 1978, 1979, O'Reilly and Mielke 1980, Brockwell *et al.* 1982), approximate P-values are based on the Pearson type III distribution which compensates for the fact that the underlying permutation distribution is often substantially skewed (Harter 1969; Mielke *et al.* 1981). In particular, the standardized test statistic given by

$$T = (\delta_0 - \mu_\delta) / \sigma_\delta$$

is presumed to follow the Pearson type III distribution with the density function given by

$$f(y) = \frac{(-2\gamma)^{4/\gamma^2}}{\Gamma(4/\gamma^2)} [-(2 + y\gamma) / \gamma]^{(4-\gamma^2)/\gamma^2} e^{-2(2+y\gamma)\gamma^2}$$

where $-\infty < y < -2/\gamma$. If $T_0 = (\delta_0 - \mu_\delta) / \sigma_\delta$ and $\gamma = \gamma_\delta < -0.001$, then

$$P(\delta < \delta_0) = \int_{-\infty}^{T_0} f(y) dy$$

is the approximate P-value (an approximate P-value based on the standard normal distribution is reported if $\gamma_\delta > -0.001$). The approximate P-value is evaluated with Simpson's rule over the interval $(T_0 - 9, T_0)$.

Incidentally, the expressions given by $\sum_{i < j} \Delta_{i,j} \Phi_i(\omega_j)$ are the primary group random variables of tests based on statistic δ . Let the symmetric distance functions be restricted to

$$\Delta_{i,j} = |x_i - x_j|^\nu$$

where $\nu > 0$. Also suppose $x_{1,n} \leq x_{2,n} \leq \dots \leq x_{n,n}$ denote the n order statistics associated with the response measurements for a specific group of size n . Then the group random variable in question here is given by $\sum_{i < j} |x_{i,n} - x_{j,n}|^\nu$ where $\sum_{i < j}$ is the sum overall i and j such that $1 < i < j < n$. If $\nu = 1$, then

$$\sum_{i=1}^{n-1} |n - 2i + 1| |x_{i,n} - \Phi| = \sum_{i < j} |x_{i,n} - x_{j,n}| \tag{2.1.5}$$

where both equation (2.1.5) is satisfied and the lefthand side of equation (2.1.5) is minimized when Φ is the

median of $x_{1,n}, \dots, x_{n,n}$. Therefore tests based on $v = 1$ are termed median based permutation techniques. Similarly, if $v = 2$,

$$\frac{1}{n} \sum_{i=1}^n (x_{i,n} - \Phi)^2 = \sum_{i < j} (x_{i,n} - x_{j,n})^2 \quad (2.1.6)$$

where both equation (2.1.6) is satisfied and the lefthand side of equation (2.1.6) is minimized when Φ is the mean of $x_{1,n}, \dots, x_{n,n}$. As a consequence, tests based on $v = 2$ are termed mean based permutation techniques.

2.2. Relationship between MRPP and other well known methods

We first describe the relationship between the permutation version of one-way analysis of variance (two sample t test when $g = 2$) and MRPP. Let $F = MS_A/MS_W$ (mean square among groups/mean square within groups) be the ordinary one-way analysis of variance statistic. If $g \geq 2$, $v = 2$, $r = 1$, $A = \sum_{i=1}^g x_{i1}$, $B = \sum_{i=1}^g x_{i2}$, $N = K$ and $C_i = (n_i - 1)/(N - g)$ for $i = 1, \dots, g$, then the identity relating F and δ is given by

$$N \delta = 2 (NB - A^2) / ([N - g + (g - 1) F]).$$

Because F is based on $v = 2$, the previously mentioned geometry problem of the underlying analysis space is a relevant concern for the permutation version of one-way analysis of variance. In Section 4 we will see by a series of examples how this geometry problem is crucial in determining P -values under certain circumstances.

The relation between simple two-sample linear rank tests and MRPP is now described. Let $H = \sum_{i=1}^N f_i U_i$ be a two-sample linear rank test statistic where $U_i = 1$ if ω_i belongs to S_1 and 0 otherwise, $n_1 = n$, $n_2 = N - n$, $g = 2$, $r = 1$, f_i is a score function of the rank order value of x_{i1} from below relative to the finite population of N response measurements, $A = \sum_{i=1}^N f_i$, $B = \sum_{i=1}^N f_i^2$ and $C_i = (n_i - 1)/(N - 2)$ for $i = 1$ and 2. Here the identity relating H and δ is given by

$$N(N - 2) \delta = 2 (NB - A^2 (NH - nA)^2 / [n(N - n)]).$$

Since H is also based on $v = 2$ and the geometry problem of underlying analysis space is again a relevant concern.

Whaley (1983) demonstrates the equivalence of (a) the multidimensional runs statistic developed by Friedman and Rafsky (1979); (b) the spatial autocorrelation statistic introduced by Cliff and Ord (1973); and (c) the special case of the MRPP statistic when $n_1 = n$, $n_2 = N - n$, $g = 2$, $C_i = n_i(n_i - 1) / [n_1(n_1 - 1) + n_2(n_2 - 1)]$ for $i = 1$ and 2, and $\Delta_{i,j}$ is 1 if ω_i and ω_j are linked and 0 otherwise. When $n_1 = n_2$, this interesting observation by

Whaley (1983) suggests that the performance of this statistic might be improved if the present choice of C_i is replaced with $C_i = n_i/N$ for $i = 1$ and 2 (the simple structure of $\Delta_{i,j}$ eliminates the geometry problem). As indicated by Mielke (1986), these tests usually involve a nonmetric analysis space.

2.3. Methodological description of MRBP

The techniques of this section involve recently developed analogs of MRPP which utilize blocking (Mielke and Berry 1982; Mielke and Iyer 1982). This section is followed by a description of the relationship between MRBP and well known statistical techniques that involve randomized block designs.

Let b blocks and g treatments be associated with a randomized block design. Let $(x_{1ij}, \dots, x_{rij})$ denote r commensurate response measurements corresponding to treatment i and block j (the response measurements might again be functions of response measurements or residuals adjusted by predictors). The modified MRPP statistic for this situation is given by

$$\delta = [g \binom{b}{2}]^{-1} \sum_{i=1}^g \sum_{j < k} \Delta(x_{ij}, x_{ik})$$

where $\Delta(x, y)$ is a symmetric distance function value of the points $x' = (x_1, \dots, x_r)$ and $y' = (y_1, \dots, y_r)$ in the r -dimensional Euclidean space. The underlying permutation distribution of δ (the null hypothesis) assigns equal probabilities to be

$$M = (g!)^b$$

possible allocations of the g r -dimensional response measurements to the g treatment positions within each of the b blocks. The mean, variance and skewness of δ under the null hypothesis are again denoted by μ_δ , σ_δ^2 and γ_δ respectively (Mielke, 1984). Except for subsequently described special cases when δ is equivalent to well known statistics, little is presently known about the asymptotic distribution of δ . However, under the null hypothesis and fairly reasonable conditions, it is conjectured that: (1) $g^{1/2}(\delta - \mu_\delta)$ is asymptotically a normal random variable when $b \geq 2$ is fixed and $g \rightarrow \infty$; and (2) $b(\delta - \mu_\delta)$ is asymptotically a nondegenerate and non-normal random variable with $\gamma_\delta < 0$ when $g \geq 2$ is fixed and $b \rightarrow \infty$.

The symmetric distance function is again confined to

$$\Delta(x, y) = \left(\sum_{h=1}^r |x_h - y_h|^p \right)^{1/p} \quad (2.3.1)$$

where $p > 1$ and $v > 0$. Since the choice of the symmetric distance function defines the structure of the underlying

analysis space of these procedures, the discussion in Section 2.1 concerning this choice is equally pertinent here.

In a manner analogous to MRPP, small values of δ imply a concentration of the response measurements associated with each of the g treatments (i.e., over blocks). Therefore $P(\delta < \delta_0)$ is again the P-value associated with δ_0 (the realized value of δ). Though an efficient algorithm for calculating the exact P-value for an observed value of δ exists, this approach becomes prohibitively expensive when M is large (e.g. greater than 10^6). Noting that $b=6$ and $g=4$ yields $M=1.9 \times 10^8$ or that $b=4$ and $g=6$ yields $M=2.7 \times 10^{11}$, the necessity for an approximation technique is obvious for even relatively small randomized block configurations.

As in Section 2.1, approximate P-values are again based on the Pearson type III distribution to compensate for the commonly encountered substantial skewness of the underlying permutation distribution of δ . Thus the standardized test statistic given by

$$T = (\delta - \mu_\delta) / \sigma_\delta$$

is again presumed to follow the Pearson type III distribution and the approximate P-value is obtained by the previously described approach in Section 2.1.

For certain response patterns involving randomized block designs, the statistic δ mentioned above is not able to detect treatment differences. Such situations can occur when the block differences exceed the magnitude of the treatment differences. As a simple example, consider the following univariate data set with $b=2$ and $g=3$.

	Treatment		
	1	2	3
Block 1	1	2	3
Block 2	5	6	7

If $v=1$, then $\delta=4$ for all $M=36$ possible allocations. It is therefore impossible to detect treatment differences. This disturbing behavior of δ can be rectified by aligning the responses within each block. The importance of alignment was initially pointed out by Hodges and Lehmann (1962). Alignment is accomplished for the present example by replacing x_{ij} by $x_{ij} - x_j^*$ where x_j^* is the median of x_{1j}, \dots, x_{gj} for $j=1, \dots, b$. Similarly x_j^* could just as well have been the mean of the responses in the j th block. The statistic δ is then computed after the alignment. After alignment with the present example, the following data set is obtained.

Treatment	1	2	3
Block 1	-1	0	1
Block 2	-1	0	1

Now the value of δ is 0 while the random variable δ assumes the values 0, 2/3 and 4/3 with the respective probabilities 1/6, 1/3 and 1/2 under the null hypothesis.

A general recommendation is that a test associated with δ should be accomplished after alignment. Also note that if $v=2$ and $r=1$, then the inferential results based on the random variable δ remain unaffected by the alignment. This latter property holds because the F statistic does not depend on the values of x_1^*, \dots, x_b^* .

2.4. Relationship between MRBP and other randomized block statistical methods.

The relation between δ and the classical F statistic for testing the null hypothesis of a randomized block design is initially described. If $v=2$ and $r=1$, then the functional relation between F and δ is given by

$$F = (b-1) [2SS_T - g(b-1)\delta] / [g(b-1)\delta - 2SS_B]$$

where the corrected total sum of squares is given by $SS_T = (\sum_{i=1}^g \sum_{j=1}^b x_{ij}^2) - SS_M$, the block sum of squares is given by $SS_B = \{ \sum_{j=1}^b [(\sum_{i=1}^g x_{ij})^2 / g] \} - SS_M$, and $SS_M = (\sum_{i=1}^g \sum_{j=1}^b x_{ij})^2 / bg$. Thus F and δ are equivalent under the null hypothesis since SS_T and SS_B are invariant relative to the $(g!)^b$ permutations of the response measurements. (For this and other cases involving univariate responses, the response measurement subscript is omitted, i.e., $x_{1ij} = x_{ij}$.) Incidentally, δ is equivalent to Cochran's Q test statistic if $r=1$ and each x_{ij} is either 0 or 1.

Let R denote the ordinary Pearson correlation coefficient. If $v=2$, $b=2$ and $r=1$, then the functional relation between R and δ is given by

$$R = (\mu_\delta - \delta) / (2S_1 S_2)$$

where $R = [\sum_{i=1}^g (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)] / (gS_1 S_2)$, $\mu_\delta = S_1^2 + S_2^2 + (\bar{x}_1 - \bar{x}_2)^2$, $\bar{x}_j = (\sum_{i=1}^g x_{ij}) / g$ and $S_j^2 = [\sum_{i=1}^g (x_{ij} - \bar{x}_j)^2] / g$ for $j=1$ and 2. Then R and δ are equivalent under the null hypothesis because $\bar{x}_1, \bar{x}_2, S_1$ and S_2 are invariant relative to the $(g!)^2$ response measurement permutations.

If $g=2$, $r=1$, $x_{1j} = -x_{2j} = x_j$ and $|x_j| > 0$ for $j=1, \dots, b$, then the test based on δ is equivalent to an extended class of permutation techniques for matched pairs (Mielke and Berry 1982).

If $v = 2, r = 1$ and the response measurements for each block are replaced by their corresponding ranks, then the test based on δ is equivalent to the Friedman two-way analysis of variance (Kendall coefficient of concordance). The values of $\mu_\delta, \sigma_\delta^2$ and γ_δ for this case are

$$\mu_\delta = (g^2 - 1) / 6, \quad \sigma_\delta^2 = [(g + 1)(g^2 - 1)] / [18b(b - 1)],$$

and

$$\gamma_\delta = -\{8(b - 2)^2 / [(g - 1)b(b - 1)]\}^{1/2}.$$

Furthermore $\rho = 1 - \delta / \mu_\delta$ is Spearman's rho (a measure of correlation) when $b = 2$. In the present context, the 'correlation' measure $\rho = 1 - \delta / \mu_\delta$ can be interpreted in a much broader setting.

If $v = r = 1$ and the response measurements for each block are again replaced by their corresponding ranks, then the test based on δ is the Euclidean space analog of the Friedman two-way analysis of variance (Kendall coefficient of concordance). The values of $\mu_\delta, \sigma_\delta^2$ and γ_δ for this case are

$$\mu_\delta = (g^2 - 1) / (3g), \quad \sigma_\delta^2 = \frac{2(g + 1)(2g^2 + 7)}{45 g^2 b(b - 1)}$$

and

$$\gamma_\delta = \frac{(g + 2)(2g^2 + 31) \theta(g) + (8g^4 + 29g^2 + 71)}{[49(g + 1)(2g^2 + 7)]^3} \cdot \frac{(b - 2)/(g - 1)}{b(b - 1)/40]^{1/2}}$$

where $\theta(g) = 0$ or 1 if $g = 2$ or $g \geq 3$, respectively. In this case δ is the Spearman footrule statistic when $b = 2$ (Diaconis and Graham 1977).

3. Implementation of the method

As mentioned in Section 2, for moderately large sample sizes ($N > 24$) the Pearson type III distribution is used to estimate P values. For that purpose, it is necessary to compute values for $\mu_\delta, \sigma_\delta^2$ and γ_δ (mean, variance and skewness). In Sections 3.1 and 3.2 we outline efficient algorithms to estimate $\mu_\delta, \sigma_\delta^2$ and γ_δ for both MRPP and MRBP. We also give a very brief description of the computer program designed to carry out these computations. The FORTRAN coding for the computer programs is given in Appendix I and II.

3.1. Algorithms for the MRPP statistical method

Efficient computational expressions for $\mu_\delta, \sigma_\delta^2$ and γ_δ are given by:

$$\mu_\delta = D(1),$$

$$\sigma_\delta^2 = 2 \left\{ \sum_{i=1}^g C_i^2 [n_i^{(2)}]^{-1} - [N^{(2)}]^{-1} \right\} [D(2) - 2D(2') + D(2'')]$$

$$+ 4 \left\{ \sum_{i=1}^g C_i^2 n_i^{-1} - N^{-1} \right\} [D(2') - D(2'')],$$

$$\gamma_\delta = \{E[\delta^3] - 3\mu_\delta \sigma_\delta^2 - \mu_\delta^3\} / \sigma_\delta^3,$$

and

$$E[\delta^3] = 4 \sum_{i=1}^g C_i^3 [n_i^{(2)}]^{-2} D(3) + 8 \sum_{i=1}^g C_i^3 n_i^{(3)} [n_i^{(2)}]^{-3} [3D(3') + D(3'*)]$$

$$+ 8 \sum_{i=1}^g C_i^3 n_i^{(4)} [n_i^{(2)}]^{-3} [3D(3'**) + D(3'***)]$$

$$+ 6 \sum_{i=1}^g C_i^2 \{1 - C_i + C_i n_i^{(4)} [n_i^{(2)}]^{-2}\} [n_i^{(2)}]^{-1} D(3'')$$

$$+ 12 \sum_{i=1}^g C_i^2 \{(1 - C_i) n_i^{(3)} + C_i n_i^{(6)} [n_i^{(2)}]^{-1}\} [n_i^{(2)}]^{-2} D(3''')$$

$$+ \sum_{i=1}^g C_i \{(1 - C_i)(1 - 2C_i) + 3C_i(1 - C_i) n_i^{(4)}\} [n_i^{(2)}]^{-2}$$

$$+ C_i^2 n_i^{(6)} [n_i^{(2)}]^{-3} D(3''')$$

where $n_i^{(c)} = n! / (n - c)!$.

The $D(i)$ where $i = 1, 2, 2', 2'', 3, 3', 3'', 3''', 3''', 3^*, 3^{**},$ and 3^{***} are calculated with the use of the following efficient algorithm: let

$$d_{kj} = \sum_{j=1}^N \Delta_{j,j'}^k \text{ and } d_k = \sum_{j=1}^N d_{kj}$$

for $k = 1, 2$ and 3 , where $\Delta_{j,j'}$ is calculated as in 2.1.3.

Then

$$D(1) = \frac{1}{N^{(2)}} d_1, D(2) = \frac{1}{N^{(2)}} d_2, D(2') = \frac{1}{N^{(3)}}$$

$$[\sum_{j=1}^N d_{jj}^2 - d_2],$$

$$D(2'') = \frac{1}{N^{(4)}} [d_1^2 - 4N^{(3)} D(2') - 2d_2], D(3) =$$

$$\frac{1}{N^{(2)}} d_3,$$

$$D(3') = \frac{1}{N^{(3)}} [\sum_{j=1}^N d_{1j} d_{2j} - d_3], D(3'') = \frac{1}{N^{(4)}}$$

$$[d_1 d_2 - 4N^{(3)} D(3') - 2d_3],$$

$$D(3^*) = \frac{6}{N^{(3)} J_1 < J_2 < J_3} \Delta_{J_1, J_2} \Delta_{J_1, J_3} \Delta_{J_2, J_3},$$

$$D(3^{**}) = \frac{1}{N^{(4)}} [2 \sum_{j_1 < j_2} \Delta_{j_1, j_2} d_{1j_1} d_{1j_2} - 2N^{(3)}$$

$$D(3') - N^{(3)} D(3^*) - d_3],$$

$$D(3^{***}) = \frac{1}{N^{(4)}} [\sum_{j=1}^N d_{jj}^3 - 3N^{(3)} D(3') - d_3],$$

$$D(3''') = \frac{1}{N^{(6)}} [N^{(3)} d_1 D(2') - 4N^{(4)} D(3^{**}) -$$

$$2N^{(4)} D(3^{***}) - 4N^{(3)} D(3') - 2N^{(3)} D(3^*)],$$

and

$$D(3''''') = \frac{1}{N^{(6)}} [N^{(4)} d_1 D(2'') - 8N^{(5)} D(3''') -$$

$$4N^{(4)} D(3'') - 8N^{(4)} D(3^{**})]$$

where $\sum_{j_1 < j_2} ()$ denotes the summation over distinct integral values of J_1 and J_2 ranging from 1 through N with J_1 less than J_2 etc.

The computer program (see Appendix I for the detailed FORTRAN coding) is designed to implement the above algorithms. Program DRIVER reads the required parameters and the variable format, calls subroutine MRPP,

and prints the results of the MRPP analysis. Subroutine MRPP reads in the raw data, labels each grouping with a group number, calculates the test statistic (DELTA), the expected value of DELTA (MEAN), the variance of DELTA (SIGMA2), the skewness of DELTA (GAMMA 1), and then calls subroutine TYPE3 to obtain the P-value based on the Pearson type III distribution. Subroutine TYPE3 computes and returns the P-value based on the Pearson type III distribution using numerical integration involving the Newton-Cotes formula for a tenth order polynomial.

The program for MRPP in Appendix I has the following restrictions and constants.

LIMIT = the smallest power (-675.00) without an underflow condition occurring in function FOX3.

N = the number of partitions (500) over which the integration takes place in subroutine TYPE3. This value may be modified for increased/decreased accuracy/efficiency.

NINE = the number of standard deviations (9) over which the integration takes place in subroutine TYPE3.

TEST = selects the normal distribution if the absolute value of GAMMA1 is less than or equal to 0.001, otherwise, the Pearson Type III distribution is selected.

3.2 Algorithm for the MRBP statistical method

The computational expressions for $\mu_\delta, \sigma_\delta^2$ and γ_δ in MRBP are given by

$$\mu_\delta = [g^2 \binom{b}{2}]^{-1} \sum_{r(s)} \text{SIJ}(r, s),$$

$$\sigma_\delta^2 = [g^2 \binom{b}{2}]^{-2} \frac{1}{g-1} \sum_{r(s)} \{ [\text{SIJ}(r, s)]^2 - g \text{TIJ2}(r, s) + g^2 \text{SIJ2}(r, s) \},$$

$$K_3(\delta) = [g^2 \binom{b}{2}]^{-3} \frac{1}{g-1} \left(\frac{1}{g-2} \sum_{r(s)} \{ 4[\text{SIJ}(r, s)]^3 - 6g \text{SIJ}(r, s) \text{TIJ2}(r, s) + 6g^2 \text{UIJ}(r, s) + 2g^2 \text{TIJ3}(r, s) + 3g^2 \text{SIJ}(r, s) \text{SIJ2}(r, s) - 3g^2 \text{VI}(r, s) + g^4 \text{SIJ3}(r, s) \} \right.$$

$$\left. - \frac{6}{g-1} \sum_{r(s)(t)} \{ \text{SIJ}(r, s) \text{SIJ}(r, t) \text{SIJ}(s, t) \} \right.$$

$$\left. - g \{ \text{SIJ}(s, t) \text{WI}(r, s, t) + \text{SIJ}(r, t) \text{WI}(s, r, t) \} \right.$$

$$\left. + \text{SIJ}(r, s) \text{WI}(t, r, s) \right\} + g^2 \text{YIJ}(r, s, t)$$

$$- g^3 \text{ZIJK}(r, s, t)),$$

and

$$\gamma_{\delta} = K_3 (\delta) / \sigma_{\delta}^3$$

Let's define d_{ij} as a multivariate observation in treatment i block j and $\Delta(i, r; j, s) = \Delta(d_{ir}, d_{js})$ where $\Delta(x, y)$ is calculated as in 2.3.1. The quantities involved in the estimation of $\mu_{\delta}, \sigma_{\delta}^2$, and γ_{δ} are calculated in the following way:

$$SJ(i, r, s) = \sum_{j=1}^g \Delta(i, r; j, s),$$

$$SJ2(i, r, s) = \sum_{j=1}^g [\Delta(i, r; j, s)]^2,$$

$$SJ3(i, r, s) = \sum_{j=1}^g [\Delta(i, r; j, s)]^3,$$

$$SIJ(r, s) = \sum_{i=1}^g SJ(i, r, s),$$

$$SIJ2(r, s) = \sum_{i=1}^g SJ2(i, r, s),$$

$$SIJ3(r, s) = \sum_{i=1}^g SJ3(i, r, s),$$

$$TJ2(i, r, s) = [SJ(i, r, s)]^2 + [SJ(i, s, r)]^2,$$

$$TJ3(i, r, s) = [SJ(i, r, s)]^3 + [SJ(i, s, r)]^3,$$

$$TIJ2(r, s) = \sum_{i=1}^g TJ2(i, r, s),$$

$$TIJ3(r, s) = \sum_{i=1}^g TJ3(i, r, s),$$

$$U(i, r; j, s) = \Delta(i, r; j, s) SJ(i, r, s) SJ(j, s, r),$$

$$UJ(i, r, s) = \sum_{j=1}^g U(i, r; j, s),$$

$$UIJ(r, s) = \sum_{i=1}^g UJ(i, r, s),$$

$$V(i, r, s) = SJ(i, r, s) SJ2(i, r, s) + SJ(i, s, r) SJ2(i, s, r),$$

$$VI(r, s) = \sum_{i=1}^g V(i, r, s),$$

$$W(i; r, s, t) = SJ(i, r, s) SJ(i, r, t),$$

$$WI(r, s, t) = \sum_{i=1}^g W(i; r, s, t),$$

$$Y(i, j; r, s, t) = \Delta(i, r; j, s) SJ(i, r, t) SJ(j, s, t)$$

$$+ \Delta(i, r; j, t) SJ(i, r, s) SJ(j, t, s)$$

$$+ \Delta(i, s; j, t) SJ(i, s, r) SJ(j, t, r),$$

$$YJ(i; r, s, t) = \sum_{j=1}^g Y(i, j; r, s, t),$$

$$YIJ(r, s, t) = \sum_{i=1}^g YJ(i; r, s, t),$$

$$Z(i, j, k; r, s, t) = \Delta(i, r; j, s) \Delta(i, r; k, t) \Delta(j, s; k, t),$$

$$ZK(i, j; r, s, t) = \sum_{k=1}^g Z(i, j, k; r, s, t),$$

$$ZJK(i; r, s, t) = \sum_{k=1}^g ZK(i, j; r, s, t),$$

and

$$ZIJK(r, s, t) = \sum_{i=1}^g ZJK(i; r, s, t).$$

The computer program designed to implement this algorithm and calculate P consist of 3 main subroutines (for detail FORTRAN coding of MRBP see Appendix II). Subroutine MRBP reads in the raw data and calculates the distance functions, the test statistic, the mean, and variance, and the skewness of the sampling distribution. Acting as a driver program, subroutine MRBP then calls subroutines MRPP and PVALUE to calculate the probability value. Subroutine MRBP sequentially reads in the x_{irk} raw scores ($i = 1, \dots, g; r = 1, \dots, b; k = 1, \dots, c$) calculates the required distance functions, and writes the computed values onto a random-access file for later retrieval. This subroutine also calculates δ_0 (as in 2.3.1), $\mu_{\delta}, \sigma_{\delta}$, and γ_{δ} (as outlined above) and ϕ (here $\phi = 1 - \delta/\mu_{\delta}$ is a generalized measure of agreement). It then computes the standardized test statistics $T_0 = (\delta_0 - \mu_{\delta})/\sigma_{\delta}$.

4. Examples

This section is devoted to show properties of MRPP and MRBP on different types of data sets (both artificial and experimentally derived). Comparisons are made between MRPP or MRBP and standard statistical method. A special emphasis is given to the relationships among the data space geometry, analysis space geometry, and derived P-values.

4.1. Example 1

The first example (Table 3) shows a univariate case of two groups (S_1 and S_2) with equal means but different point concentrations (similar to the case of Fig. 2). This example can not be treated with a standard parametric test (let's say a one- or two-sided t test) because the variances are not homogeneous and thus violates one of the basic assumptions. But, even if the variances were homogeneous, a parametric test would be worthless because it only deals with location shifts in the mean and in this case the means are equal. Furthermore, even if the

Table 3. Two groups S_1 and S_2 with equal mean but different degree of point concentration.

Group	Data points											Mean
S_1	5	6	4	5.5	6.2	5	5.2	5.1	4.9	5	5.2	5.19
S_2	15	12	0.1	0.1	15	0.5	0.2	0.42	0.1	13.5	0.17	5.19

means were different and the variances homogeneous, a mean based test would be of little value since the means of S_2 (5.19 in this case) have very little meaning as an interpretation tool because the points are concentrated in the extremes (in a sense, the mean stands in an empty space with no real data point close to it). The groups, regardless of their equal means, still show a totally different data structure. S_1 is very concentrated while S_2 is dispersed. This type of data set is particularly common in a comparison between a seeded pasture and a native pasture. Both could have the same biomass but a different distribution of the biomass over the landscape. For instance, the seeded pasture is bound to be homogeneous while the native one, especially in semiarid regions, is more likely to have the vegetation concentrated on clumps with many empty spaces. The MRPP and MRBP allows us to test for that without any concern about the underlying nature of the density functions involved. As such they give us an important tool to perform vegetation analysis which are not possible with other methods adding reliability to our results because no caveats are needed about normal distributions, linearity, homogeneity of variances, etc.

There are two ways to treat the problem presented in Table 3. One way involves a standard two group test ($g = 2$), in which case we test for the probability of having one group with a concentrated data structure and another with a dispersed one. The other way is to perform a one sample test ($g = 1$) and use the other group as an excess group (the $g + 1$ groups explained in Section 2). In this case, we will look at the probability of observing a given data point concentration when a sample of n individuals ($n = 11$ in this case) is drawn from a population of N ($N = 22$ in this case). Table 4 shows the test results. It can be seen from the results that there is great disparity between the use of $v = 1$ (both data and analysis spaces are metric spaces) and $v = 2$ (analysis space is a nonmetric space). As explained in the introduction, this is due to the fact

that congruence principle is violated with $v = 2$ since the data space is a metric space but the analysis space is a nonmetric space. Furthermore when values of Table 3 are plotted in Fig. 4 and therefore shown in a metric space, it becomes intuitively obvious that S_1 is more concentrated than S_2 . This feature is captured by the analysis based on $v = 1$ but completely missed by the analysis with $v = 2$. We showed in Sections 2.2 and 2.4 that MRPP and MRBP can be related to other standard statistical methods when $v = 2$ is used. As a result this type of distortion is bound to be found in all standard statistical tests. We will cover this geometry problem more in detail in Example 4.

4.2. Example 2

This example (Table 5) is similar to Example 1 (Table 3) except that it was constructed in such a way as to make the two groups share not only an equal mean but also an equal median as well. By so doing, this case is untreatable not only by standard parametric procedures but also

Table 4. Two-sided P-value comparisons between groups S_1 and S_2 from Table 3 using MRPP with $v = 1$ and $v = 2$. We also show a one sample MRPP for S_1 and S_2 with $v = 1$.

Test	P
MRPP S_1 vs. S_2 ($v = 1, g = 2$)	0.247×10^{-2}
MRPP S_1 vs. S_2 ($v = 2, g = 2$)	0.81
S_1 one sample MRPP ($v = 1, g = 1$)	0.905×10^{-4}
S_2 one sample MRPP ($v = 1, g = 1$)	0.99

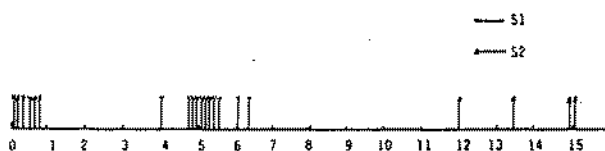


Fig. 4. Plot for values of Table 3.

Table 5. Two groups S_1 and S_2 with equal mean and median but different degrees of point concentration.

Group	Data points											Mean	Median
S_1	4.1	4.1	4.2	4.5	4.6	4	3.9	3.98	3.99	3.89	3.92	4.10	4
S_2	12	7	8	6	6	4	1	1	.01	.01	.16	4.10	4

Table 6. Two-sided P-value comparison between groups S_1 and S_2 from Table 5 using MRPP with $v = 1$ and $v = 2$.

	MRPP $v = 1$	MRPP $v = 2$
S_1 vs. S_2	0.27×10^{-2}	0.82

by most nonparametric procedures that are based on medians. Like Example 1, it is visually obvious that the values in S_1 are again far more concentrated than the values in S_2 . This structural characteristic may have very important ecological consequences. MRPP allows us to investigate this type of data and the results are shown in Table 6. Again there is an enormous discrepancy between the p-values associated with $v = 1$ as compared to $v = 2$.

4.3. Example 3

One of the most common situations in vegetation research, or almost all biological research, is the presence of outliers (one or two points that seem not to agree with the general pattern of observed values). If these values are legitimate (in the sense that they are not errors in measurement), the investigator is confronted with the problem of how to treat the data. One way is to simply eliminate the data values in question from the analysis (which often goes unreported). The other way is simply to include the values in question (either as it is or with the use of some transformation) in the analysis. This inclusion (as we will see) may lead to the acceptance of the null hypothesis even though this acceptance may be intuitively wrong. The use of MRPP or MRBP with $v = 1$ gives an easy solution to this problem without having to eliminate such values. Figure 5 displays a scatter diagram of 20 bivariate observations partitioned into two groups, S_1 and S_2 , with one outlier with coordinates $x_1 = 53$ and $x_2 = 51$ belonging to S_1 . By looking at Fig. 5, one would intuitively conclude that S_1 and S_2 are distinctively different. To test this hypothesis, we ran MRPP based on $v = 1$ and $v = 2$ and also four standard parametric procedures: Hotelling T^2 , Pillai's criterion, Roy's maximum root and Wilk's lambda. Table 7 shows the results and once again there is a large difference between MRPP with $v = 1$ and the rest. Only MRPP based on $v = 1$ agrees with the intuitive observation. The reason is that, with the use of $v = 2$ in MRPP or the least squares procedure in the other four parametric tests, square distances are involved (that is our analysis space is a nonmetric space) and therefore the influence of outliers are made paramount. In the case of MRPP based on $v = 1$, the congruence principle is upheld since the data space

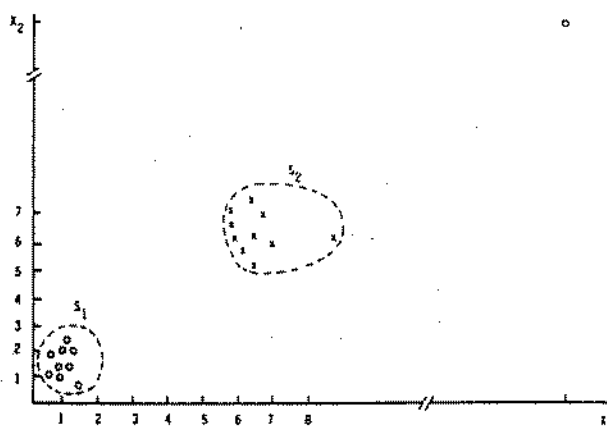


Fig. 5. Diagram showing points of two groups (S_1 and S_2) plotted as a function of 2 measured responses; $o = S_1$, $x = S_2$.

corresponds to the analysis space. The result is that one outlier does not overwhelm the remaining values and, as such, we have a better correspondence between our intuitive observations and the test results.

4.4. Example 4

The intent of this example is to further consider the influence an individual outlier, w , when the remaining values are specified. For this purpose, consider two groups S_1 and S_2 with $n_1 = n_2 = n$, $N = 2n$ and $C_1 = C_2 = 1/2$. Let $n-1$ values associated with S_1 be 0 and let the remaining value be w . Also let all n values associated with S_2 be 1. For MRPP with $v = 1$ we have (using the algorithms of Section 3.1)

$$\delta = |w|/n,$$

$$\mu\delta = \frac{n-2}{2n-1} + \frac{w}{n} \quad \text{if } w \geq 1,$$

$$\mu\delta = \left(n - \frac{w}{n}\right) / (2n-1) \quad \text{if } 0 < w < 1,$$

Table 7. P-values for MRPP with $v = 1$ and $v = 2$, Hotelling T^2 , Pillai's criterion, Roy's maximum root and Wilk's lambda on the data shown in Fig. 5.

Test procedure	P-value
MRPP, $v = 1$	0.65×10^{-4}
MRPP, $v = 2$	0.99
Hotelling T^2	0.99
Pillai's criterion	0.99
Roy's maximum root	0.99
Wilk's lambda	0.99

$$\mu_{\delta} = \frac{n}{2n-1} - \frac{w}{n} \quad \text{if } w \leq 0,$$

and

$$\sigma_{\delta}^2 = (n+1)(n-2) / [(n-1)(2n-1)^2(2n-3)] \quad \text{if } w \geq 1,$$

$$\sigma_{\delta}^2 = \left(\frac{n-4w}{2n-3} + \frac{2w^2}{n-1} \right) / (2n-1)^2 \quad \text{if } 0 < w < 1,$$

$$\sigma_{\delta}^2 = n / [(2n-1)^2(2n-3)] \quad \text{if } w \leq 0,$$

then statistic $T = (\delta - \mu_{\delta}) / \sigma_{\delta}$ is

$$T = - [(n-1)(n-2)(2n-3) / (n+1)]^{1/2} \quad \text{if } w \geq 1,$$

$$T = - (n-2w) / \left(\frac{n-4w}{2n-3} + \frac{2w^2}{n-1} \right)^{1/2} \quad \text{if } 0 < w < 1, \quad (4.1.1)$$

$$T = - [n(2n-3)]^{1/2} \quad \text{if } w \leq 0.$$

Similarly for MRPP with $v = 2$ we have

$$\delta = w^2/n,$$

$$\mu_{\delta} = \frac{n-2w}{2n-1} + \frac{w^2}{n},$$

and

$$\sigma_{\delta}^2 = \left(\frac{n-4w}{2n-3} + \frac{2w^2}{n-1} \right) / (2n-1)^2,$$

then

$$T = - (n-2w) / \left(\frac{n-4w}{2n-3} + \frac{2w^2}{n-1} \right)^{1/2}. \quad (4.1.2)$$

If n is fixed, equation (4.1.1) shows that for the case of $v = 1$, T remains practically unchanged for all values of w . This indicates that the outlier will have a minimum influence in the statistical outcome when $v = 1$ since T is the basic statistics used to calculate the P-value (see Section 2.1). However, for the same case with $v = 2$ equation (4.1.2) implies that

$$T \rightarrow \begin{cases} + [2(n-1)]^{1/2} & \text{when } w \rightarrow +\infty \\ - [2(n-1)]^{1/2} & \text{when } w \rightarrow -\infty \end{cases}$$

Thus a statistical inference involving $v = 2$ may be either reversed by an individual value (outlier) which is inconsistent with the remaining data (when w is a large positive value in this example) or weakened by an individual value which is consistent with the remaining data (when w is a large negative value in this example). As empirical evidence to support these results, univariate and a multivariate examples follow.

The univariate data set in Table 8 involves two groups, S_1 with 14 specified values and one arbitrary value

Table 8. Frequencies of S_1 and S_2 values.

Value	S_1	S_2
58.2	1	0
58.3	4	0
58.4	5	0
58.5	3	1
58.6	1	3
58.7	0	4
58.8	0	5
58.9	0	2
X	1	0

(denoted by x) and S_2 with 15 specified values. The 14 specified values of S_1 were selected to be smaller than the values of S_2 but allow for a slight overlap. The data point x in group S_1 was allowed to take values from 18.4 to 98.4. For each value of x , the two groups were tested for statistical differences with a test which satisfies the congruence principle (MRPP with $v = 1$) and with two tests which do not satisfy the congruence principle (MRPP with $v = 2$ and the standard parametric two-sample t test). The P-values associated with each of the tests are shown in Table 9. As can be readily observed, the P-values of MRPP with $v = 1$ are only slightly affected by the different values of x while the P-values of both MRPP with $v = 2$ and the two-sample t test are highly dominated by the different values of x .

For the multivariate example, consider the data presented in Fig. 5. After altering the value of the outlier in a symmetrical way (i.e., replace $x_1 = 53$ and $x_2 = 51$ with $x_1 = -53$ and $x_2 = -51$), the same analysis shown in Table 7 is accomplished again. In this case, the P-values for Hotelling T^2 , Pilla's criterion, Roy's maximum root and Wilk's lambda drop from 0.99 to 0.19, and the P-value for MRPP based on $v = 2$ drops from 0.99 to 0.14×10^{-4} whereas the P-value for MRPP based on $v = 1$

Table 9. P-value comparisons of MRPP with $v = 1$, MRPP with $v = 2$, and the two-sample t test.

X	MRPP		Two-sample t test
	$v = 1$	$v = 2$	
18.4	6.6×10^{-7}	4.2×10^{-6}	0.27
48.4	6.6×10^{-7}	4.1×10^{-6}	0.15
53.4	6.6×10^{-7}	4.5×10^{-6}	5.7×10^{-2}
56.4	6.6×10^{-7}	5.4×10^{-6}	2.3×10^{-3}
57.4	6.6×10^{-7}	4.9×10^{-6}	1.8×10^{-5}
58.4	6.6×10^{-7}	1.9×10^{-6}	4.9×10^{-9}
59.4	9.8×10^{-6}	1.5×10^{-3}	2.0×10^{-3}
60.4	9.8×10^{-6}	0.14	0.16
63.4	9.8×10^{-6}	0.99	1.00
68.4	9.8×10^{-6}	0.99	0.62
98.4	9.7×10^{-6}	0.99	0.39

Table 10. Percent relative cover of the dominant plant species groups of each treatment on years 1 and 6 of secondary succession.

Species Group	Treatment 1		Treatment 2		Treatment 3		Treatment 4	
	Yr 1	Yr 6	Yr 1	Yr 6	Yr 1	Yr 6	Yr 1	Yr 6
Grasses	49.04	62.15	14.27	38.07	1.27	43.80	0.04	5.44
Perennial Forbs	33.90	32.86	13.39	21.05	1.14	25.31	1.48	11.79
Annual Forbs	14.63	2.54	66.12	2.29	91.56	21.04	92.62	6.57
Shrubs	2.32	2.17	5.55	37.71	5.91	9.25	5.40	76.86

Treatment 1: the vegetation was mechanically removed with minimal disturbance to topsoil (A and B horizons).

Treatment 2: the vegetation was mechanically removed and the topsoil scarified to a depth of 30 cm.

Treatment 3: topsoil and subsoil (C horizon) were removed to a depth of 1 m. The material was mixed together and replaced.

Treatment 4: two layers of 1 m of soil were removed and replaced in a reverse order with the second layer placed on the surface.

remains essentially unchanged, as anticipated by the previous results. These results are remarkable since they demonstrate that the alteration of only one data point in a symmetrical way can reverse conclusions of tests not satisfying the congruence principle (i.e., accepting to reject the null hypothesis even though the basic data structure has not been changed).

4.5. Example 5

In this section we present the application of MRPP and MRBP to real case studies in vegetation research. The first case (Table 10) comes from a study on secondary succession (Biondini *et al.* 1985). The hypothesis to test was that increased levels of soil disturbance (from Treatment 1 that consisted of mechanical removal of vegetation with minimal disturbance to A and B soil horizon to Treatment 4 which consisted of physically inverting the soil to a depth of 2 m) significantly alter the direction of secondary succession. The main patterns of vegetation succession in this study were given by the following changes through time: (1) perennial grasses (the dominant species being: *Agropyron riparium*, *A. smithii*, *Koeleria cristata*, *Oryzopsis hymenoides*, and *Stipa comata*); (2) perennial forbs (the dominant species being: *Sphaeralcea coccinea*, *Erigeron engelmannii*, *Phlox longifolia*, *Senecio multilobatus*, and *Trifolium gymnocarpon*); (3) annual forbs (the dominant species being: *Salsola iberica*); and (4) shrubs (the dominant species being: *Artemisia tridentata*, *Chrysothamnus nauseosus*, *C. viscidiflorus*, and *Gutierrezia sarothrae*).

The hypothesis was formally tested with MRPP. The percent relative cover of perennial grasses, perennial forbs, annual forbs, and shrubs shown in Table 8 were used as the multivariate observation which characterized the species composition of each treatment. Treatments were analyzed at two points in time: (1) one year into succession; and (2) six years into succession. Treatments

2, 3, and 4 were shown to be different (P -value = 0.035) from Treatment 1 one year into succession. No significant difference (P -value = 0.11) was found among Treatments 2, 3, and 4. Six years later, Treatments 1 and 3 were marginally similar (P -value = 0.090). Treatment 2 was different from Treatments 1 and 3 (P -value = 0.042) and Treatment 4 was different from Treatment 2 (P -value = 0.03). The hypothesis that increased levels of soil disturbance can alter the pattern of secondary succession was accepted because of the following: (1) the species composition of Treatments 2, 3 and 4 was not statistically different in the early stages of succession while six years later the species composition of Treatments 2, 3 and 4 was statistically different; and (2) Treatment 4 became a shrub-dominated community while Treatments 1 and 3 and to a lesser extent Treatment 2 became grass-forb dominated communities.

The second case comes from a study associated with mine reclamation research (Redente *et al.* 1982). An area that had been shallowly disturbed was seeded with a combination of grasses, forbs, and shrubs (for details see Redente *et al.* 1982) and subjected to six treatments. The treatments were: (a) no fertilizer; (b) low fertilizer; (c) high fertilizer; (d) mulch and fertilizer; (e) mulch and low fertilizer; and (f) mulch and high fertilizer (Table 9). The experiment was organized in a randomized block design with 3 blocks (Table 9). The biomasses of three shrubs (*Atriplex canescens*, *Ceratoides lanata*, and *Ephedra viridis*) were used as the multivariate observation that characterized each treatment (Table 9). The data was analyzed with the use of MRBP (based on $v = 1$) and, for comparison purposes with a series of standard multivariate parametric analyses (see Table 11). The null hypothesis (no difference between treatments) could be rejected with a P -value 0.068 with MRBP based on $v = 1$ (which satisfies the congruence principle) but would have to be accepted if one were to use any of the other standard multivariate parametric methods (which do not satisfy the congruence principle).

Table 11. Biomass data on shrubs (g/m^2) for a reclamation study on a shallowly disturbed site (Redente *et al.* 1982). The three shrubs measured on each treatment (listed from top to bottom) are: *Atriplex canescens*, *Ephedra viridis* and *Ceratoides lanata*. The treatments are (a) no fertilizer; b) low fertilizer (56 kg N/ha + 28 kg P/ha); (c) high fertilizer (112 kg N/ha + 56 kg P/ha); (d) mulch (2.2 MT/ha of wood fiber hydromulch) and no fertilizer; (e) mulch and low fertilizer; and (f) mulch and high fertilizer.

	Treatment					
	a	b	c	d	e	f
Block 1	0.33	6.67	6.33	3.83	9.67	14.50
	1.00	5.00	8.50	8.00	1.33	0.50
	2.17	2.00	2.17	1.33	3.67	2.17
Block 2	2.50	18.67	4.17	11.50	8.33	21.67
	0.83	0.17	1.67	2.50	1.50	0.83
	4.33	2.83	2.17	3.50	2.50	2.17
Block 3	1.00	8.67	5.17	0.67	23.67	7.33
	0.50	0.50	3.33	7.00	0.50	7.50
	2.33	2.00	2.00	2.50	3.33	4.50

5. Summary and conclusions

In the previous sections we explored the shortcomings inherent in most of the standard statistical parametric techniques when it comes to their application to ecological research in general and vegetation analysis in particular. With regard to the univariate (multivariate) parametric methods, a series of limitations had to do with the basic assumptions of normal (multivariate normal) distributions, homogeneity of variance (variance-covariance matrix) and linearity that have to be made about the underlying structure of the population under study. Another potentially more serious limitation, which applies not only to parametric methods but also to most nonparametric methods, is the lack of congruence between the geometry of the data space and the geometry of the analysis space. While the data space on all of the vegetation and ecological studies is a well known metric space (a Euclidean space), the analysis space for most of the traditionally used statistical methods is a complex nonmetric space. As the examples that were presented have shown, this can lead to results that are both (1) inconsistent with intuitive observations, and (2) unstable when only a single value is changed (see Example 4).

A new statistical method called multiresponse permutation procedures (MRPP) and its randomized block

designed analog (MRBP) have been presented in detail in both their theoretical as well as practical foundations. Efficient algorithms as well as the associated computer programs needed to implement MRPP and MRBP were outlined in Sections 3.1 and 3.2. The method does not require any assumptions about the underlying distribution structure of the population under study and, more importantly, allows us to utilize a metric analysis space. Furthermore, while the most commonly used parametric methods only test for shifts in the location of the means, MRPP and MRBP allow us to test for a variety of population structural differences (shifts in the location of the medians, changes in data concentration, etc.) which enlarges the scope of the hypotheses that can be accurately tested in vegetation analyses. We believe that MRPP and MRBP can be an important analytical tool that can (1) enlarge the scope of research studies in ecology and vegetation science; and (2) be performed with a strict adherence to the hypothetical-deductive portion of the scientific method.

References

- Berry, K.J. and P.W. Mielke. 1984. Computation of exact probability values for multi-response permutation procedures (MRPP). *Commun. Statist.-Simulation Comput.* 13: 417-432.
- Biondini, M.E., C.D. Bonham and E.F. Redente. 1985. Secondary successional patterns in a sagebrush (*Artemisia tridentata*) community as they relate to soil disturbance and soil biological activity. *Vegetatio* 60: 25-36.
- Brockwell, P.J., P.W. Mielke and J. Robinson. 1982. On non-normal invariance principles for multi-response permutation procedures. *Austral. J. Statist.* 24: 33-41.

Table 12. Results of MRBPP, $v = 1$ and other multivariate parametric tests run on the data shown on Table 10.

Test procedure	P-value
MRBPP based on $v = 1$	0.068
Hotelling's criteria	0.21
Pillai's criterion	0.13
Roy's maximum root	0.20
Wilk's lambda	0.15

- Cliff, A.D. and T.K. Ord. 1973. *Spatial Autocorrelation*. Pion Limited, London, England.
- Diaconis, P. and R.L. Graham. 1977. Spearman's footrule as a measure of disarray. *J. Roy. Statist. Soc. Ser. B* 39: 262-268.
- Friedman, J.H. and L.C. Rafsky. 1979. Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *Ann. Statist.* 7: 697-717.
- Harter, H.L. 1969. A new table of percentage points of the Pearson type III distribution. *Technometrics* 11: 177-187.
- Hodges, J.L. and E.L. Lehman. 1962. Rank methods for combination of independent experiments in analysis of variance. *Ann. Math. Statist.* 33: 482-497.
- Mielke, P.W. 1978. Clarification and appropriate inferences for Mantel and Valand's nonparametric multivariate analysis technique. *Biometrics* 34: 277-282.
- Mielke, P.W. 1979. On asymptotic non-normality of null distributions of MRPP statistics. *Comm. Statist. A* 8: 1541-1550. *Errata*: A 10, 1975; A 11, 847.
- Mielke, P.W. 1984. Meteorological applications of permutation techniques based on distance function. In: P.R. Krishnaiah and P.K. Sen (eds.) *Handbook of Statistics*. Vol. 4. pp. 813-830. Elsevier Science Publishers.
- Mielke, P.W. 1985. Geometric concerns pertaining to applications of statistical tests in the atmospheric sciences. *J. Atmospheric Sci.* 42: 1209-1212.
- Mielke, P.W. 1986. Non-metric statistical analyses: Some metric alternatives. *J. Statist. Plann. Inference* 13: 377-387.
- Mielke, P.W. and K.J. Berry. 1982. An extended class of permutation techniques for matched pairs. *Commun. Statist. A* 11: 1197-1207.
- Mielke, P.W., K.J. Berry and G.W. Brier. 1981. Applications of multiresponse permutation procedures for examining seasonal changes in monthly mean sea-level pressure patterns. *Mon. Wea. Rev.* 109: 120-126.
- Mielke, P.W., K.J. Berry and E.S. Johnson. 1976. Multi-response permutation procedures for a priori classifications. *Comm. Statist. A* 5: 1409-1424.
- Mielke, P.W. and H.K. Iyer. 1982. Permutation techniques for analyzing multi-response data from randomized block experiments. *Comm. Statist. A* 11: 1427-1437.
- Mielke, P.W. and P.K. Sen. 1981. On asymptotic non-normal null distributions for locally most powerful rank test statistics. *Comm. Statist. A* 10: 1079-1094.
- O'Reilly, F.J. and P.W. Mielke. 1980. Asymptotic normality of MRPP statistics from invariance principles of U-statistics. *Comm. Statist. A* 9: 629-637.
- Orlóci, L. 1979. Non-linear data structure and their description. In: L. Orlóci, C. Radhakrishna Rao and W. Stiteler (eds.) *Multivariate Methods in Ecological Work*. International Co-operative Publishing House, Fairland, Maryland, USA.
- Peter, R. 1980. From natural history to ecology. *Perspectives in Biology and Medicine*. Winter 1980: 191-203.
- Redente, E.F., T.B. Doerr, C.E. Grygiel, E. Allerdings, J.M. Stark and M.E. Biondini. 1982. Effect of plant species, soil material, and cultural practices upon plant establishment on succession. In: E.F. Redente and C.W. Cook (eds.) *Revegetation Studies on Oil Shale Related Disturbances in Colorado*. pp. 1-25. DOE/EV/0401806. Dept. Range Science, Colorado State University, Fort Collins, Colorado.
- Sala, O.E. and W.K. Lauenroth. 1982. Small rainfall events: an ecological role in semiarid regions. *Oecologia* 53: 301-304.
- Sen, P.K. 1970. The Hajek-Renyi inequality from sampling from a finite population. *Sankhya Ser. A* 32: 181-188.
- Sen, P.K. 1972. Finite population sampling and weak convergence of a Brownian bridge. *Sankhya Ser. A* 34: 85-90.
- Simberloff, D. 1981. The sick science of ecology. *Eidemia* 1(1).
- Whaley, F.S. 1983. The equivalence of the three independently derived permutation procedures for testing the homogeneity of multidimensional samples. *Biometrics* 39: 741-745.


```

DO 160 I = 1, 12
  B(1) = ZERO
160 CONTINUE
      CALCULATE THE B(I) VALUES BASED ON
      A SINGLE SUMMATION
DO 170 I = 1, NCASES
  B(1) = B(1) + DC(1, I)
  B(2) = B(2) + DC(2, I)
  B(3) = B(3) + DC(3, I) ** 2
  B(4) = B(4) + DC(4, I)
  B(5) = B(5) + DC(5, I)
  B(6) = B(6) + DC(6, I) * DC(1, I)
  B(10) = B(10) + DC(1, I) ** 3
170 CONTINUE
      CALCULATE THE B(I) VALUES BASED ON
      A DOUBLE SUMMATION
DO 180 I = 1, NCASES - 1
  DO 180 J = I + 1, NCASES
    B(9) = B(9) + DC(1, J) * DC(1, I)
180 CONTINUE
190 CONTINUE
      B(9) = B(9) * 2.0
      CALCULATE THE B(I) VALUES BASED ON
      A TRIPLE SUMMATION
DO 240 I = 1, NCASES - 2
  DO 230 J = I + 1, NCASES - 1
    DO 220 K = J + 1, NCASES
      SUM = DEL(I, J) * DEL(I, K) * DEL(I, K)
      B(7) = B(7) + SUM
220 CONTINUE
230 CONTINUE
240 CONTINUE
      B(7) = B(7) * 6.0
      CALCULATE THE REMAINING B(I) VALUES
B(4) = B(1) ** 2
B(8) = B(1) * B(2)
B(11) = B(1) * B(3)
B(12) = B(1) ** 3
      CALCULATE THE VALUES FOR A1 TO A11
A1 = B(1)
A2 = B(2)
A3 = B(3) - B(2)
A4 = B(4) - TWO * A2 - FOUR * A3
A5 = B(5)
A6 = B(6) - B(5)
A7 = B(7)
A8 = B(8) + TWO * B(5) - TWO * B(6)
A9 = B(9) + B(5) - TWO * B(6) - B(7)
A10 = B(10) + TWO * B(5) - THREE * B(6)
A11 = B(11) - TWO * B(10) - B(7) + TWO * B(5) + B(9) + TEN
      * B(6) - B(8)
A12 = B(12) - FOUR * A3 + TWO * A7 + A10 + THREE * A6 + A9
      - SIX * A8 + TWO * A11
      CALCULATE THE VALUES FOR K-SUPER(I)
      WITH I = 1, ..., 6
M(1) = NCASES
DO 280 I = 1, 6
  M(I) = M(I - 1) * (NCASES - I + 1)
280 CONTINUE
      CALCULATE THE VALUES FOR D1 TO D12
D1 = A1 / M(2)
D2 = A2 / M(2)
D3 = A3 / M(2)
D4 = A4 / M(4)
D5 = A5 / M(2)
D6 = A6 / M(3)
D7 = A7 / M(3)
D8 = A8 / M(6)
D9 = A9 / M(6)
D10 = A10 / M(6)
D11 = A11 / M(5)
D12 = A12 / M(6)
      CALCULATE THE VALUES FOR K-SUPER(I)-
      SUPER(J) WITH I = 1, ..., 6, M(I) AND
      J = 1, ..., 6
DO 300 I = 1, 6, M(I)
  NI(1, I) = I * M(I)
  DO 290 J = 1, 6
    NI(I, J) = NI(I, J - 1) * (I * M(I) - J + 1)
290 CONTINUE
300 CONTINUE
      INITIALIZE THE SUB-TOTAL REGISTERS
      S(I), I = 1, ..., 8
DO 310 I = 1, 8
  S(I) = ZERO
310 CONTINUE
      CALCULATE THE SUB-TOTALS FOR THE
      S(I) VALUES WITH I = 1, ..., 8
DO 340 I = 1, 8, M(I)
  S(1) = S(1) + C(I) ** 2 / M(I, 2)
  S(2) = S(2) + C(I) ** 2 / M(I, 1)
  S(3) = S(3) + C(I) ** 3 / M(I, 2) ** 2
  S(4) = S(4) + C(I) ** 3 * M(I, 2) / M(I, 2) ** 3
  S(5) = S(5) + C(I) ** 2 * ((ONE - C(I)) / M(I, 2) + C(I) *
1  M(I, 4) / M(I, 2) ** 3)
  S(6) = S(6) + C(I) ** 3 * M(I, 4) / M(I, 2) ** 3
  S(7) = S(7) + C(I) ** 2 * ((ONE - C(I)) * M(I, 3) /
1  M(I, 2) ** 2 + C(I) * M(I, 5) / M(I, 2) ** 3)
  S(8) = S(8) + C(I) * ((ONE - C(I)) * (ONE - TWO * C(I)) +
1  THREE * C(I) * (ONE - C(I)) * M(I, 4) / M(I, 2) ** 2
2  + C(I) ** 2 * M(I, 6) / M(I, 2) ** 3)
340 CONTINUE
      CALCULATE THE EXPECTED VALUE OF THE
      TEST STATISTIC (DELTA)
MEAN = D1
      CALCULATE THE VARIANCE OF DELTA

```

NRPP 237
NRPP 238
NRPP 239
NRPP 240
NRPP 241
NRPP 242
NRPP 243
NRPP 244
NRPP 245
NRPP 246
NRPP 247
NRPP 248
NRPP 249
NRPP 250
NRPP 251
NRPP 252
NRPP 253
NRPP 254
NRPP 255
NRPP 256
NRPP 257
NRPP 258
NRPP 259
NRPP 260
NRPP 261
NRPP 262
NRPP 263
NRPP 264
NRPP 265
NRPP 266
NRPP 267
NRPP 268
NRPP 269
NRPP 270
NRPP 271
NRPP 272
NRPP 273
NRPP 274
NRPP 275
NRPP 276
NRPP 277
NRPP 278
NRPP 279
NRPP 280
NRPP 281
NRPP 282
NRPP 283
NRPP 284
NRPP 285
NRPP 286
NRPP 287
NRPP 288
NRPP 289
NRPP 290
NRPP 291
NRPP 292
NRPP 293
NRPP 294
NRPP 295
NRPP 296
NRPP 297
NRPP 298
NRPP 299
NRPP 300
NRPP 301
NRPP 302
NRPP 303
NRPP 304
NRPP 305
NRPP 306
NRPP 307
NRPP 308
NRPP 309
NRPP 310
NRPP 311
NRPP 312
NRPP 313
NRPP 314
NRPP 315
NRPP 316
NRPP 317
NRPP 318
NRPP 319
NRPP 320
NRPP 321
NRPP 322
NRPP 323
NRPP 324
NRPP 325
NRPP 326
NRPP 327
NRPP 328
NRPP 329
NRPP 330
NRPP 331
NRPP 332
NRPP 333
NRPP 334
NRPP 335
NRPP 336
NRPP 337
NRPP 338
NRPP 339
NRPP 340
NRPP 341
NRPP 342
NRPP 343
NRPP 344
NRPP 345
NRPP 346
NRPP 347
NRPP 348
NRPP 349
NRPP 350
NRPP 351
NRPP 352
NRPP 353
NRPP 354
NRPP 355
NRPP 356
NRPP 357
NRPP 358
NRPP 359
NRPP 360
NRPP 361
NRPP 362
NRPP 363
NRPP 364
NRPP 365

```

SIGMA2 = TWO * (S(1) - ONE / M(2)) / (M(2) - TWO * B(3) + B(4))
FOUR * (S(2) - ONE / M(1)) / (M(1) - B(4))
      CALCULATE THE SKEWNESS OF DELTA
DELTA = FOUR * S(3) * B(5) * EIGHT * N(4) * (THREE * B(6) + B(7) +
1  SIX * S(5) * B(1) * EIGHT * N(4) * (THREE * B(6) + B(7) +
2  12.0 * S(7) * B(1) * N(4) * B(12))
KAPPA3 = DELTA - THREE * MEAN * SIGMA2 - MEAN ** 3
GAMMA1 = KAPPA3 / SIGMA2 ** 3.5
      CALL TYPE3 (DELTA, MEAN, SIGMA2, GAMMA1, PVALUE, IFAULT)
RETURN
END
SUBROUTINE TYPE3 (DELTA, MEAN, SIGMA2, GAMMA1, PVALUE, IFAULT)
      DESCRIPTION OF VARIABLES IN CALL LIST:
DELTA = INPUT, OBSERVED (REALIZED) VALUE OF DELTA
MEAN = INPUT, EXPECTED (MEAN) VALUE OF DELTA
SIGMA2 = INPUT, VARIANCE OF THE DELTA DISTRIBUTION
GAMMA1 = INPUT, SKEWNESS OF THE DELTA DISTRIBUTION
PVALUE = OUTPUT, PROBABILITY OF A REALIZED DELTA VALUE THIS
      EXTREME OR MORE EXTREME
IFAUULT = OUTPUT, ZERO IF NO ERRORS
USAGE:
SUBROUTINE TYPE3 COMPUTES AND RETURNS A PROBABILITY VALUE
      BASED ON THE PEARSON TYPE III PROBABILITY DISTRIBUTION.
      MEAN, MEAN, NINE, LIMIT
PARAMETER (N = 500, LIMIT = 0.9999, TEST = 0.001)
DATA C1 / 1000.0, C2 / 100000.0, C3 / 50000.0, C4 / 22200.0,
1  C5 / 200000.0, C6 / 222000.0, C7 / 32200.0, C8 / 20000.0,
2  ZERO / 0.00, ONE / 1.00, TWO / 2.00, THREE / 3.00,
3  FOUR / 4.00, FIVE / 5.00, SIX / 6.00, EIGHT / 8.00,
4  NINE / 9.00
      TEST FOR THE NUMERICAL VALUE OF THE
      GAMMA1 VALUE AND THE STANDARDIZED
      TEST STATISTIC
T = (DELTA - MEAN) / SQRT (SIGMA2)
IF (ABS (GAMMA1) .GT. TEST) THEN
  ALPHA = TWO / ABS (GAMMA1)
  IF (GAMMA1 .LT. ZERO .AND. T .GT. LIMIT * ALPHA) THEN
    PVALUE = ONE
  ELSE IF (GAMMA1 .GT. ZERO .AND. T .LT. -LIMIT * ALPHA) THEN
    PVALUE = ZERO
  ELSE
    GO TO 10
  END IF
ELSE
  CALCULATE A PROBABILITY VALUE BASED
  ON THE NORMAL DISTRIBUTION
DATA D0 / 0.2316419007, D1 / 0.3193815307, D2 / -0.3585637827,
1  D3 / 1.781477927, D4 / -1.821255978, D5 / 1.330274429 /
2  D = ONE / (ONE + D0 * ABS (T))
3  P1 = FOUR * ATAN (ONE)
4  PVALUE = (EXP (-ONE * (T ** 2 + TWO)) / SQRT (TWO * P1)) *
1  D * (D1 + D * (D2 * D + (D3 * D + (D4 * D + D5)))
2  END IF
      CHECK ON VALUE OF IFAULT
IF (PVALUE .LT. ZERO) IFAULT = SEVEN
RETURN
      CALCULATE THE VALUE OF THE NATURAL
      LOG OF 8
LOG 8 = (HAT (ALPHA ** 2) * ALPHA ** 2 * (ONE - LOG (ALPHA)))
      BEGIN THE INTEGRATION ROUTINE
THETA = T
IF (GAMMA1 .LE. ZERO) THETA = THETA + NINE
C = (T - THETA) / (HAT (N))
Y0 = C1 + FOUR * (ALPHA, GAMMA1, B, THETA + (1 - 1) * C) - C2
Y1 = C1 + FOUR * (ALPHA, GAMMA1, B, THETA + (1 - 2) * C) + C3 + FOUR *
1  C4 + FOUR * (ALPHA, GAMMA1, B, THETA + (1 - 3) * C) - C5 + FOUR *
2  C6 + GAMMA1, B, THETA + (1 - 4) * C) + C7 + FOUR * (ALPHA, GAMMA1,
3  B, THETA + (1 - 5) * C) + C8 + FOUR * (ALPHA, GAMMA1, B,
4  THETA + (1 - 6) * C) + C9 + FOUR * (ALPHA, GAMMA1, B,
5  THETA + (1 - 7) * C) + C10 + FOUR * (ALPHA, GAMMA1, B,
6  THETA + (1 - 8) * C)
70 CONTINUE
      CALCULATE A PROBABILITY VALUE BASED
      ON A PEARSON TYPE III DISTRIBUTION
PVALUE = FIVE * (T - THETA) * (Y0 - Y1 + Y2) / (C8 * N)
      CHECK ON VALUE OF IFAULT
IF (PVALUE .LT. ZERO) IFAULT = EIGHT
RETURN
END
FUNCTION GHAT (Y)
      DESCRIPTION OF VARIABLES IN CALL LIST:
Y = INPUT, VARIABLE PASSED FROM SUBROUTINE TYPE3
USAGE:
FUNCTION GHAT CALCULATES STIRLING'S APPROXIMATION FOR THE
      NATURAL LOG OF (UPPER-CASE) GAMMA (Y).
DATA C1 / 12.0, C2 / 260.0, C3 / 1260.0, C4 / 1680.0,
1  C5 / 1188.0, ZERO / 0.00, ONE / 1.00, TWO / 2.00,
2  FOUR / 4.00, FIVE / 10.0, HALF / 0.50

```

NRPP 366
NRPP 367
NRPP 368
NRPP 369
NRPP 370
NRPP 371
NRPP 372
NRPP 373
NRPP 374
NRPP 375
NRPP 376
NRPP 377
NRPP 378
NRPP 379
NRPP 380
NRPP 381
NRPP 382
NRPP 383
NRPP 384
NRPP 385
NRPP 386
NRPP 387
NRPP 388
NRPP 389
NRPP 390
NRPP 391
NRPP 392
NRPP 393
NRPP 394
NRPP 395
NRPP 396
NRPP 397
NRPP 398
NRPP 399
NRPP 400
NRPP 401
NRPP 402
NRPP 403
NRPP 404
NRPP 405
NRPP 406
NRPP 407
NRPP 408
NRPP 409
NRPP 410
NRPP 411
NRPP 412
NRPP 413
NRPP 414
NRPP 415
NRPP 416
NRPP 417
NRPP 418
NRPP 419
NRPP 420
NRPP 421
NRPP 422
NRPP 423
NRPP 424
NRPP 425
NRPP 426
NRPP 427
NRPP 428
NRPP 429
NRPP 430
NRPP 431
NRPP 432
NRPP 433
NRPP 434
NRPP 435
NRPP 436
NRPP 437
NRPP 438
NRPP 439
NRPP 440
NRPP 441
NRPP 442
NRPP 443
NRPP 444
NRPP 445
NRPP 446
NRPP 447
NRPP 448
NRPP 449
NRPP 450
NRPP 451
NRPP 452
NRPP 453
NRPP 454
NRPP 455
NRPP 456
NRPP 457
NRPP 458
NRPP 459
NRPP 460
NRPP 461
NRPP 462
NRPP 463
NRPP 464
NRPP 465
NRPP 466
NRPP 467
NRPP 468
NRPP 469
NRPP 470
NRPP 471
NRPP 472
NRPP 473
NRPP 474
NRPP 475
NRPP 476
NRPP 477
NRPP 478
NRPP 479
NRPP 480
NRPP 481
NRPP 482
NRPP 483
NRPP 484
NRPP 485
NRPP 486
NRPP 487
NRPP 488
NRPP 489
NRPP 490
NRPP 491
NRPP 492
NRPP 493
NRPP 494
NRPP 495

```

C
PI = FOUR * ATAN (ONE)
SUM = ZERO
N = T * TEN
DO 10 I=1, 10
  SUM = SIN + LOG (X - 1)
10 CONTINUE
CHAT = (X + HALF) * LOG (X) - X + HALF * LOG (TWO * PI) + ONE
1 / (C1 * X) + ONE / (C2 * X ** 2) + ONE / (C3 * X ** 5)
2 ONE / (C4 * X ** 7) + ONE / (C5 * X ** 9) + SUM
RETURN
END

C
FUNCTION FOX3 (ALPHA, GAMMA1, B, X)
  DESCRIPTION OF VARIABLES IN CALL LIST:
  ALPHA - INPUT, PARAMETER (ALPHA) PASSED FROM TYPES
  BETA - INPUT, PARAMETER (BETA) PASSED FROM TYPES
  GAMMA - INPUT, PARAMETER (GAMMA) PASSED FROM TYPES
  HRPP 496 C # - INPUT, PARAMETER (B) PASSED FROM TYPES
  HRPP 497 C X - INPUT, VARIABLE PASSED FROM SUBROUTINE TYPES
  HRPP 498 C
  HRPP 499 C USAGE:
  HRPP 500 C
  HRPP 501 C FUNCTION FOX3 CALCULATES F(X) FOR THE PEARSON TYPE III
  HRPP 502 C PROBABILITY DISTRIBUTION.
  HRPP 503 C
  REAL LIMIT
  PARAMETER (LIMIT = +.75, 0)
  DATA ZERO /0.00/, ONE /1.00/
  HRPP 504
  HRPP 505
  HRPP 506
  HRPP 507 C
  HRPP 508
  HRPP 509
  HRPP 510
  HRPP 511
  HRPP 512
  HRPP 513
  HRPP 514
  HRPP 515
  HRPP 516
  HRPP 517
  HRPP 518
  HRPP 519
  HRPP 520
  HRPP 521
  HRPP 522
  HRPP 523
  HRPP 524
  HRPP 525
  HRPP 526
  HRPP 527
  HRPP 528
  HRPP 529
  HRPP 530
  HRPP 531
  HRPP 532
  HRPP 533
  HRPP 534
  HRPP 535
  HRPP 536
  HRPP 537
  HRPP 538

```

Appendix II. FORTRAN program for the MRBP statistical method. For details about the algorithms used see Section (3.2).

```

PROGRAM MRBP
IMPLICIT REAL*(A-H,O-Z)

C
THIS FORTRAN PROGRAM COMPUTES THE TEST STATISTIC AND ASSOCIATED
P-VALUE FOR AN ANALYSIS OF A RANDOMIZED BLOCK EXPERIMENT (MRBP).
THE CORRESPONDENCE BETWEEN A RANDOMIZED BLOCK EXPERIMENT AND A
RANDOMIZED BLOCK EXPERIMENT CAN BE USED TO GET A CORRELATION
COEFFICIENT AS WELL. THE MAXIMUM VALUES OF C, B AND R CAN BE
CHANGED FOR ANY EXAMPLE. THE PRESENT MAXIMUM VALUES OF C, B
AND R IN THIS PROGRAM ARE RESPECTIVELY 10, B AND 5.
C
COMMON /BL1/D(60,60)
COMMON /BL2/SJ(10,8,8),SJ2(10,8,8),SJ3(10,8,8),UJ(10,8,8)
COMMON /BL3/V,C,B,R,DATA(10,8,5)
COMMON /BL4/VI(8,8,8),YIJ(8,8,8),ZIK(8,8,8)
COMMON /BL5/SIJ(8,8),SIJ2(8,8),SIJ3(8,8),UIJ(8,8)
1 TIJ2(8,8),TIJ3(8,8),VI(8,8)
COMMON /BL6/T,CAN,PROB
CHARACTER FMT#60,HEAD#60,FTH#64,FOUR#64
INTEGER C,B,R

C
THE DATA MATRIX MUST BE IN THE FOLLOWING SEQUENCE WITH EACH
OBJECT'S R RESPONSE VALUES ON A SEPARATE LINE AS FOLLOWS:
A(1,1,1),A(1,1,2),....,A(1,1,R)
A(1,2,1),A(1,2,2),....,A(1,2,R)
.....
A(1,B,1),A(1,B,2),....,A(1,B,R)
A(2,1,1),A(2,1,2),....,A(2,1,R)
.....
A(2,B,1),A(2,B,2),....,A(2,B,R)
.....
A(C,1,1),A(C,1,2),....,A(C,1,R)
.....
A(C,B,1),A(C,B,2),....,A(C,B,R)

C
THE INPUT DATA ARE IN FREE FORMAT. SPECIFICALLY, V IS DISTANCE EXPONENT,
C IS # OF GROUPS, B IS # OF BLOCKS, AND R IS # OF RESPONSES.

WRITE(*,500)
500 FORMAT(' ANALYSIS OF RANDOMIZED BLOCK EXPERIMENT (MRBP) ///
1 - ENTER NAME OF INPUT DATA FILE: \')
READ(*,*)FTH
WRITE(*,501)
501 FORMAT(' ENTER NAME OF OUTPUT/DEVICE FILE: \')
READ(*,*)FOUR
OPEN(5,FILE=FTH)
OPEN(6,FILE=FOUR,STATUS='NEW')
READ(5,*)C,B,R
READ(5,*)V,C,B,R
WRITE(6,600) HEAD,V,R,C,B
600 FORMAT (' ANALYSIS OF RANDOMIZED BLOCK EXPERIMENT (MRBP): '
1 4X,A/4X,'DISTANCE EXPONENT: ',F4.1/4X,'WITH ',15,
2 ' RESPONSES, ',15,' GROUPS, ',15,' BLOCKS.',
3 //IX,'** INPUT DATA: ')

DO 110 I=1,C
DO 107 J=1,B
READ(5,*)END=50) (DATA(I,J,K),K=1,R)
WRITE(6,558) I,J,(DATA(I,J,K),K=1,R)
107 CONTINUE
110 CONTINUE
58 FORMAT (215.8F10.2)
CLOSE(5)
CALL MRPP
GO TO 200
59 WRITE(6,60)
60 FORMAT (' ERROR--END OF FILE ENCOUNTERED')
WRITE(6,61) C,B,R
61 FORMAT (' CHECK THESE: C = ',15,' B = ',15,' R = ',15)
200 STOP
END

C
SUBROUTINE MRPP
IMPLICIT REAL*(A-H,O-Z)
COMMON /BL1/D(60,60)
COMMON /BL2/SJ(10,8,8),SJ2(10,8,8),SJ3(10,8,8),UJ(10,8,8)
COMMON /BL3/V,C,B,R,DATA(10,8,5)
COMMON /BL4/VI(8,8,8),YIJ(8,8,8),ZIK(8,8,8)
COMMON /BL5/SIJ(8,8),SIJ2(8,8),SIJ3(8,8),UIJ(8,8)
1 TIJ2(8,8),TIJ3(8,8),VI(8,8)
COMMON /BL6/T,CAN,PROB
INTEGER C,B,R

C
ONE=1.000
ZERO=0.000
X=ONE+V/2
BCZ=ONE+B*(B-1)/2
DO 10 I=1,C
DO 9 J=1,B
DO 8 K=1,C
LO=1
IF (I.EQ.K) LO=J
DO 7 L=LO,B
IJ=B*(I-1)+J
KL=B*(K-1)+L
D(IJ,KL)=ZERO
DO 6 M=1,R
D(IJ,KL)=D(IJ,KL)+(DATA(I,J,M)-DATA(K,L,M))**2
6 CONTINUE
D(IJ,KL)=D(IJ,KL)**X
D(KL,IJ)=D(IJ,KL)
7 CONTINUE
8 CONTINUE
9 CONTINUE
DO 30 IS=1,B
DO 29 IR=1,B
IF (IR.EQ.IS) GO TO 28
SIJ(IR,IS)=ZERO
SIJ2(IR,IS)=ZERO
SIJ3(IR,IS)=ZERO
DO 27 I=1,C
SJ(I,IR,IS)=ZERO
SJ2(I,IR,IS)=ZERO
SJ3(I,IR,IS)=ZERO
DO 26 J=1,C
IRR=(I-1)*B+IR
JSS=(J-1)*B+IS
SJ(I,IR,IS)=SJ(I,IR,IS)+D(IRR,JSS)
SJ2(I,IR,IS)=SJ2(I,IR,IS)+D(IRR,JSS)**2
SJ3(I,IR,IS)=SJ3(I,IR,IS)+D(IRR,JSS)**3
26 CONTINUE
SIJ(IR,IS)=SIJ(IR,IS)+SJ(I,IR,IS)
SIJ2(IR,IS)=SIJ2(IR,IS)+SJ2(I,IR,IS)
SIJ3(IR,IS)=SIJ3(IR,IS)+SJ3(I,IR,IS)
27 CONTINUE
28 CONTINUE
29 CONTINUE
30 CONTINUE
T2=ZERO
IF (B.LE.2) GO TO 51
DO 50 IT=3,B
IT1=IT-1
DO 49 IS=2,IT1
IS1=IS-1
DO 48 IR=1,IS
W1(IR,IS,IT)=ZERO
W1(IS,IR,IT)=ZERO
W1(IT,IR,IS)=ZERO
Y1(IR,IS,IT)=ZERO
Z1(K,IR,IS,IT)=ZERO
DO 47 I=1,C
W1(IR,IS,IT)=W1(IR,IS,IT)+SJ(I,IR,IS)+SJ(I,IR,IT)
W1(IS,IR,IT)=W1(IS,IR,IT)+SJ(I,IR,IS)+SJ(I,IS,IT)
W1(IT,IR,IS)=W1(IT,IR,IS)+SJ(I,IT,IR)+SJ(I,IT,IS)
DO 46 J=1,C
IRR=B*(I-1)+IR
JSS=B*(J-1)+IS
JTT=B*(J-1)+IT
ISS=B*(I-1)+IS
YIJ(IR,IS,IT)=YIJ(IR,IS,IT)+
D(IRR,JSS)*S1(I,IR,IT)+S2(J,IS,IT)+
D(IRR,JTT)*S1(I,IR,IS)+S2(J,IT,IS)+
D(ISS,JTT)*S1(I,IS,IR)+S2(J,IT,IR)
DO 45 K=1,C
KTT=B*(K-1)+IT
ZIK(IR,IS,IT)=ZIK(IR,IS,IT)+
D(IRR,JSS)*D(IRR,KTT)+D(JSS,KTT)
45 CONTINUE
46 CONTINUE
47 CONTINUE
T2=T2+SIJ(IR,IS)+SIJ(IR,IT)+SIJ(IS,IT)+
SIJ(IS,IT)+W1(IR,IS,IT)+C-
SIJ(IR,IT)+W1(IS,IR,IT)+C-

```

```

3          S1J(IR,IS)=W1(IT,IR,IS)*C
4          Y1J(IR,IS,IT)*C-C-Z1JK(IR,IS,IT)*C*C
48         CONTINUE
49         CONTINUE
50         CONTINUE
51         T2=T2*(C-1)
52         CONTINUE
DO 100 IS=2,B
  IS1=IS-1
  DO 99 IR=1,IS1
    TIJ2(IR,IS)=ZERO
    TIJ3(IR,IS)=ZERO
    VI(IR,IS)=ZERO
    U1J(IR,IS)=ZERO
    DO 98 I=1,C
      TIJ2(IR,IS)=TIJ2(IR,IS)+S1J(I,IR,IS)*2+
        S1(I,IR,IS)*2
      TIJ3(IR,IS)=TIJ3(IR,IS)+S1J(I,IR,IS)*3+
        S1(I,IS,IR)*3
      VI(IR,IS)=VI(IR,IS)+S1J(I,IR,IS)+S1J(I,IR,IS)+
        S1(I,IS,IR)*S1J(I,IS,IR)
      UJ(I,IR,IS)=ZERO
      DO 97 J=1,C
        IRR=B-(I-1)+IR
        JSS=B-(J-1)+IS
        UJ(I,IR,IS)=UJ(I,IR,IS)+
          D(IRR,JSS)*S1J(I,IR,IS)*S1J(J,IS,IR)
      97         CONTINUE
    U1J(IR,IS)=U1J(IR,IS)+UJ(I,IR,IS)
  99         CONTINUE
100        CONTINUE
          T1=ZERO
          EDEL=ZERO
          VAR=ZERO
          DO 200 IS=2,B
            IS1=IS-1
            DO 199 IR=1,IS1
              IF (C.LE.2) GO TO 198
              T1=T1+(S1J(IR,IS)**3)+S1J(IR,IS)*TIJ2(IR,IS)*C*C+
                U1J(IR,IS)*C*C*(C-1)+S1J(IR,IS)*C*C*(C-2)+
                S1J(IR,IS)*S1J2(IR,IS)*C*(C-3)+
                VI(IR,IS)*C*C*(C-3)+S1J3(IR,IS)*C*(C-3)*C
            198          EDEL=EDEL+S1J(IR,IS)
            VAR=VAR+S1J(IR,IS)*S1J(IR,IS)-TIJ2(IR,IS)*C+S1J2(IR,IS)*C*C
          199          CONTINUE
200          CONTINUE
          IF (C.LE.2) GO TO 206
          T1=T1/(C-2)
206          C1=ONE/(C2*C*C)
          C2=C1*C1
          C3=C2*C1
          EDEL=C1/EDEL
          VAR=VAR/C2/(C-1)
          GAM=C3*(T1-T2)/(C-1)
          GAM=GAM/DSQRT(VAR**3)
          DELTA=ZERO
          DO 300 IS=2,B
            IS1=IS-1
            DO 299 IR=1,IS1
              DO 298 I=1,C
                IRR=(I-1)*B+IR
                ISS=(I-1)*B+IS
                DELTA=DELTA+D(IRR,ISS)
            299          CONTINUE
          300          CONTINUE
          CD=BC2*C
          DELTA=DELTA/CD
          T=(DELTA-EDEL)/DSQRT(VAR)
          RHO=ONE-(DELTA/EDEL)
          WRITE (6,500)
500          FORMAT (/// 'RESULTS OF MRSP ANALYSIS: ')
          WRITE (6,600) DELTA
600          FORMAT (11X, 'DELTA(D)', T29, ' ', C16, 7)
          WRITE (6,610) EDEL
610          FORMAT (11X, 'EXPECTED DELTA', T29, ' ', C16, 7)
          WRITE (6,620) VAR
620          FORMAT (11X, 'VARIANCE OF DELTA', T29, ' ', C16, 7)
          WRITE (6,630) GAM
630          FORMAT (11X, 'SKEWNESS OF DELTA', T29, ' ', C16, 7)
          WRITE (6,640) RHO
640          FORMAT (11X, 'AGREEMENT MEASURE', T29, ' ', C16, 7)
          WRITE (6,650) T
650          FORMAT (11X, '(D-E(D))/STDEV(D)', T29, ' ', C16, 7)
          CALL PVALUE
          WRITE (6,670) PROB
670          FORMAT (11X, 'P-VALUE', T29, ' ', C16, 7)
          RETURN
          END
C
SUBROUTINE PVALUE
C
SUBROUTINE PVALUE CALCULATES THE PROBABILITY OF A
VALUE OF T BEING LESS THAN OR EQUAL TO THE OBSERVED
VALUE OF T
C
IMPLICIT REAL*(A-H,O-Z)
COMMON /BL6/7,CAM,PROB
C
ONE = 1.000
ZERO = 0.000
PI = DATAN(ONE)*4
IF (DABS(CAM).LT.0.0100) GO TO 50
R = ONE-2/DABS(CAM)
D = R*R
F = R*R+10
DO 10 I = 1,9
  D = D*(R*R+1)
10 CONTINUE
U = (F**2-1)*DLOG(F)/2-F*DLOG(PI**2)/2-DLOG(D)+ONE/(F**12)
& = -ONE/(F**F**360)
A = R*R-1
B = R*R*(DLOG(R)-1)-U
C1 = 0.04520
C2 = 0.0600
C3 = 0.01520
H1 = ZERO
H2 = ZERO
V = -1.9900/CAM
IF (CAM.LT.ZERO) GO TO 30
IF (T.LT.W) THEN
  PROB = ZERO
  GO TO 60
END IF
X = T
Y = T+B
DO 20 I = 1,99
  H1 = H1+DEXP(A*DLOG(R*X-C1*(2*I-1)))*R*(X+C1*(2*I-1))*B)
  H2 = H2+DEXP(A*DLOG(R*X-C2*I))*R*(X+C2*I)*B)
20 CONTINUE
H1H = DEXP(A*DLOG(R*X-C1*199))*R*(X+C1*199)*B)
H2H = DEXP(A*DLOG(R*X)-R*X*B)
H3 = DEXP(A*DLOG(R*Y)-R*Y*B)
PROB = ONE-C3*(H2H*(H1+H1H)*4+H2H*2+H3)
GO TO 60
30 IF (T.GT.W) THEN
  PROB = ONE
  GO TO 60
END IF
X = T-B
Y = T
DO 40 I = 1,99
  H1 = H1+DEXP(A*DLOG(R*X-C1*(2*I-1)))*R*(X+C1*(2*I-1))*B)
  H2 = H2+DEXP(A*DLOG(R*X-C2*I))*R*(X+C2*I)*B)
40 CONTINUE
H1H = DEXP(A*DLOG(R*X-C1*199))*R*(X+C1*199)*B)
H2H = DEXP(A*DLOG(R*X)-R*X*B)
H3 = DEXP(A*DLOG(R*Y)-R*Y*B)
PROB = C3*(H2H*(H1+H1H)*4+H2H*2+H3)
GO TO 60
50 E1 = 0.3193815300
E2 = -0.35856376200
E3 = 1.78147793700
F4 = -1.82126597800
H5 = 1.33027442900
H = 0.231641900
W = ONE/(1+DABS(T)+1)
PROB = (((E5+W*E4)*W+E3)*W+E2)*W+E1)*W+DEXP(-T**2)/DSQRT(PI**2)
IF (T.GT.ZERO) PROB = ONE-PROB
60 RETURN
END

```