

Multi-response permutation procedure as an alternative to the analysis of variance: An SPSS implementation

LI CAI

University of North Carolina, Chapel Hill, North Carolina

A permutation test typically requires fewer assumptions than does a comparable parametric counterpart. The multi-response permutation procedure (MRPP) is a class of multivariate permutation tests of group difference useful for the analysis of experimental data. However, psychologists seldom make use of the MRPP in data analysis, in part because the MRPP is not implemented in popular statistical packages that psychologists use. A set of SPSS macros implementing the MRPP test is provided in this article. The use of the macros is illustrated by analyzing example data sets.

Unlike classical tests that rely on the t , F , and χ^2 distributions and parametric assumptions to make inferences, permutation tests represent an ideal situation in which one can derive the exact probability associated with a test statistic when some or all of the parametric assumptions are relaxed (Mielke & Berry, 2001, p. 1). In many cases, the only assumption made by a permutation test is the exchangeability of the observational units under the null hypothesis (see, e.g., Hayes, 1996, for an excellent discussion of this issue). Exchangeability is a much weaker assumption than the parametric assumptions of (multi)normality, homogeneity of error variances (or homogeneity of error covariance matrices), and random sampling. Extended discussions of the merits of permutation tests can be found in book-length treatments elsewhere (e.g., Edgington, 1995; Good, 1993). Here, I will give only brief comments as to why the permutation test should be the preferred method of inference for experimental data when the parametric assumptions fail to hold.

First, as Edgington (1966) and many others have pointed out, psychological experiments rarely, if ever, involve random sampling of subjects from an infinite population. Yet psychologists seem not to be bothered by the fact that all of the classical parametric tests require the assumption of random sampling. When only the random assignment of subjects into treatment conditions is present, parametric tests can still be used, even if there is no random sampling, but then they serve purely as approximations to the more appropriate permutation tests (see, e.g., Edgington, 1966; Fisher, 1935). If the lack of computing facilities 50 or 60 years ago prevented psychological researchers from fully utilizing permutation tests, today, with high-speed computing widely available at the fingertips of every psy-

chologist, there is very little reason not to be aware of these tests.

Second, as Bradley (1977) and Micceri (1989) have pointed out, normality is a rarity in psychological data, which calls the parametric assumption of (multi)normality into question. Regarding the issue of normality, a reviewer of the present article pointed out that the assumption of normality is not directly imposed on the observed data and that the extent to which the normality assumption holds in practice can be hard to assess, so the observations of Bradley and Micceri perhaps do not render direct support for favoring permutation tests. In addition, when the sample size is sufficiently large, one can often appeal to the central limit theorem to claim the asymptotic normality of sampling distributions for the statistics of interest. Furthermore, it is often the case that parametric tests are quite robust against violations of the normality assumption. All of the above are valid arguments justifying the use of parametric tests in the presence of nonnormality. However, there do exist alternative procedures, with the permutation test being a prominent example, that do not make the normality assumption, and the use of which needs no justification even under nonnormality.

Third, classical procedures often assume homogeneity of error variances or homogeneity of error covariance matrices in order to obtain simplified distributional properties of the test statistics. Although it is generally believed that classical tests are robust against violations of the homogeneity assumption (Box, 1954), more recent results (e.g., Wilcox, 1997) indicate that the amount of heterogeneity of error variances encountered in psychological research is large enough to warrant caution when test results are interpreted, even in balanced one-way designs. In comparison, homogeneity of error variances is not an issue in the multi-response permutation procedure (MRPP) that I am introducing in this article (Mielke & Berry, 1994).

Fourth, it is generally believed that parametric tests have more power than do their nonparametric counterparts. Yet rather surprisingly, in a simulation study in

Correspondence concerning this article should be addressed to L. Cai, Department of Psychology, University of North Carolina, CB 3270, Chapel Hill, NC 27599-3270 (e-mail: cai@unc.edu).

which the power of the MRPP test was compared with that of the traditional MANOVA tests for detecting location shifts in bivariate outcomes, the permutation test had an equal amount of or more power than did the parametric tests (see, e.g., Mielke & Berry, 2001, pp. 59–64). Of course, more research is needed before one can make any conclusions about the superiority of one method over the other, but it is fair to say that parametric tests are not overwhelmingly more powerful than permutation tests in the multivariate setting. Of course, the permutation test is by no means a panacea for all problems encountered in data analysis, but at least psychological researchers should be informed of the alternative procedures that can serve as the backup to classical methods.

The MRPP was first introduced by Mielke, Berry, and Johnson (1976) as a technique for detecting the difference between a priori classified groups. It turned out to be an extremely versatile data-analytic framework from which a number of applications were derived, such as the measurement of agreement, multivariate correlation and association coefficients, and the detection of autocorrelation (see Mielke & Berry, 2001, for a complete coverage of applications of the MRPP framework).

MRPP is often analogous to widely used parametric tests. Specifically, under certain conditions that will be made explicit later in Equations 11 and 12, the MRPP procedure yields test statistics that are one-to-one transformations of the two-sample t or the ANOVA F (Mielke & Berry, 1994). With functionally equivalent test statistics, the parametric tests go on to use Student's t or Snedecor's F as a reference distribution to obtain p values, whereas the MRPP test uses the permutation distribution to obtain the p value. In MANOVA settings, a similar one-to-one mapping exists between the MRPP test statistic and the Bartlett–Nanda–Pillai (BNP) trace criterion (Mielke & Berry, 1994). Therefore, under conditions that ensure the functional equivalence between the MRPP test statistic and the traditional parametric test statistics, the MRPP can be effectively labeled a permutation version of the ANOVA, and the null hypothesis in an MRPP test is, therefore, the same as the one in a parametric test. Under those conditions, the only difference between the MRPP test and the parametric test lies in the choice of the reference distribution. The use of the permutation distribution makes MRPP a procedure with fewer assumptions and one that is more robust against violations of parametric assumptions.

Despite its advantages over traditional methods, MRPP is not widely used in psychology. I suspect that part of the reason is the lack of implementation of this method in commercial statistical software packages. Although FORTRAN programs were available from the inventors of this method (e.g., Mielke & Berry, 2001), none of the popular statistical packages offers an option to perform the MRPP tests.¹ Perhaps this explains why MRPP is relatively unknown to many psychologists. A citation index search shows that practicing psychological researchers are not citing the seminal papers on the MRPP technique, an indication that they are not using this method. In this article, I amend the situation by providing a set of SPSS macros that can be invoked just as easily as any other SPSS rou-

tines, such as the familiar GLM or ONEWAY procedures, when multivariate data arising from experimental studies are analyzed.

The remainder of this article is organized in the following manner. First, I shall briefly introduce the MRPP framework and the many analysis opportunities using MRPP. Next, I shall describe the SPSS macros in some detail. Finally, I shall analyze real data sets, using the macro to demonstrate its user friendliness.

Overview of MRPP

I shall start by presenting the basic formulations of MRPP for a simple one-way layout. Only the minimum amount of details are included. A full description of MRPP can be found in Mielke (1984).

A basic formulation. It is assumed that there are k groups (or treatments) in a fixed-effects one-way design.² The outcome measures are possibly multivariate. The total sample size is N , and each group is of size n_i , where $i = 1, \dots, k$, so that $\sum_{i=1}^k n_i = N$. Let $\mathbf{y}_I = (y_{1I}, \dots, y_{rI})'$ denote an $r \times 1$ vector of (possibly noncommensurate) multivariate response observations from subject I , where $I = 1, \dots, N$. Finally, let S_i denote a set that contains all the \mathbf{y}_I s belonging to the i th group, where $i = 1, \dots, k$.

If the multivariate responses are not commensurate—that is, not measured in the same unit—they must be standardized to the same metric before any MRPP analyses can proceed. If r is 1, the experimental design is simply a univariate one-way layout, and commensuration is not needed; but if r is greater than 1, commensuration becomes a must. For example, when children's physical growth is studied, the multivariate outcome variables might consist of measures of height and weight that are not measured in compatible units. I will elaborate on the issue of commensuration later. For the moment, it is assumed that the outcome measures are indeed commensurate.

The basic building block of an MRPP statistic is a distance function between two commensurate multivariate response vectors, \mathbf{y}_I and \mathbf{y}_J :

$$\Delta_{I,J} = \left[\sum_{h=1}^r (y_{hI} - y_{hJ})^2 \right]^{v/2}. \quad (1)$$

Equation 1 is a special case of the Minkowski family of distance functions often used in multidimensional scaling (see, e.g., MacCallum, 1988). The two most commonly used distance functions are (1) the nonmetric squared Euclidean distance,

$$\Delta_{I,J} = \sum_{h=1}^r (y_{hI} - y_{hJ})^2, \quad (2)$$

obtained by setting $v = 2$ in Equation 1, and (2) the metric Euclidean distance,

$$\Delta_{I,J} = \left[\sum_{h=1}^r (y_{hI} - y_{hJ})^2 \right]^{1/2}, \quad (3)$$

obtained by setting $v = 1$ in Equation 1. Clearly, both the Euclidean and the squared Euclidean distance functions can be written compactly in matrix notation as

$$\Delta_{I,J} = \left[(\mathbf{y}_I - \mathbf{y}_J)' (\mathbf{y}_I - \mathbf{y}_J) \right]^{v/2}. \quad (4)$$

When $v = 1$, MRPP tests based on the metric Euclidean distance has the nice property of being more robust and less influenced by outliers (Mielke & Berry, 1994). The use of the nonmetric squared Euclidean distance leads to a less robust test. However, $v = 2$ often yields easier interpretations of the test results, because many popular tests, such as the t test, the ANOVA F , and Hotelling's T^2 , implicitly involve the use of squared distance.

The MRPP statistic is a weighted average of within-group distances, where the weights are determined by the group sizes.³ Geometrically, a smaller value of the MRPP statistic indicates higher within-group concentration and larger between-group differences (Mielke, 1984). Therefore, in terms of detecting a treatment effect or between-group difference, a smaller value of the MRPP statistic is necessarily "better."

The MRPP statistic is given by

$$\delta_{\text{obs}} = \sum_{i=1}^k C_i \xi_i, \quad (5)$$

where C_i is the group weight for the i th group, such that $\sum_{i=1}^k C_i = 1$ and

$$\xi_i = \binom{n_i}{2}^{-1} \sum_{I < J} \Delta_{I,J} \Psi_{S_i}(\mathbf{y}_I) \Psi_{S_i}(\mathbf{y}_J) \quad (6)$$

is the average within-group distance for the i th group. The summation is taken over all integer values of I and J such that $1 \leq I < J \leq N$, and $\Psi_{S_i}(\mathbf{y}_I)$ is an indicator function given by

$$\Psi_{S_i}(\mathbf{y}_I) = \begin{cases} 1, & \text{if } \mathbf{y}_I \in S_i, \\ 0, & \text{otherwise.} \end{cases}$$

Mielke (1984) gives an extensive discussion on the choice of group weights. For Euclidean distance, $C_i = n_i/N$ should be used in general, and for squared Euclidean distance, $C_i = (n_i - 1)/(N - k)$ should be used.

The test of significance of δ_{obs} is carried out by assuming the null hypothesis that the

$$M = \frac{N!}{\prod_{i=1}^k n_i!},$$

possible ways of allocating the N responses into the k groups are equally probable. In other words, the null hypothesis states that the outcome observations are independent of group membership. Since each permutation yields a realized value of δ , the significance level of δ_{obs} is determined by counting the number of times the realized δ s are smaller than or equal to δ_{obs} . Hence, the p value associated with δ_{obs} is $\Pr(\delta_{\text{obs}}) = \{\#\delta \leq \delta_{\text{obs}}\}/M$.

Because M is usually a very large number, even for relatively small sample sizes, the exact null distribution of the MRPP statistic is difficult to obtain. Therefore, Mielke et al. (1976) proposed the method of using a Pearson Type III curve to approximate the discrete MRPP null dis-

tribution. This method needs only the first three moments of the MRPP null distribution. The exact mean, variance, and skewness of the null distribution of the MRPP statistic are defined as

$$\begin{aligned} \mu_{\delta} &= M^{-1} \sum_{i=1}^M \delta_i, \\ \sigma_{\delta}^2 &= M^{-1} \sum_{i=1}^M \delta_i^2 - \mu_{\delta}^2, \\ \gamma_{\delta} &= \frac{M^{-1} \sum_{i=1}^M \delta_i^3 - 3\mu_{\delta} \sigma_{\delta}^2 - \mu_{\delta}^3}{\sigma_{\delta}^3}, \end{aligned}$$

where the summation runs over all M possible permutations.

Efficient computational formulae for finding these moments are given in Mielke et al. (1976), which do not require running over all possible permutations. A standardized statistic T can then be calculated as

$$T = \frac{\delta_{\text{obs}} - \mu_{\delta}}{\sigma_{\delta}}, \quad (7)$$

and because T is a one-to-one function of δ_{obs} , the p value associated with δ_{obs} is approximated easily by integrating the Pearson Type III density function,

$$\Pr(\delta_{\text{obs}}) = \Pr(T) \doteq \int_{-\infty}^T f(x | \gamma_{\delta}) dx, \quad (8)$$

where $f(x|\gamma_{\delta})$ is the Pearson Type III density function, completely characterized by a skewness parameter (see Johnson, Kotz, & Balakrishnan, 1994, for details of the Type III curve).

Percentage points of the Pearson Type III distribution have been extensively tabulated (Harter, 1967), but these tables are rarely known to psychologists. Fortunately, it turns out that there is a direct correspondence between the χ^2 distribution and the Pearson Type III distribution. Because most statistical packages, including SPSS, have a built-in cumulative distribution function (CDF) of the χ^2 distribution, the approximate p value for the MRPP statistic becomes easy to obtain. According to Harter, when the skewness parameter is positive, the Pearson Type III distribution is directly related to a standardized central χ^2 distribution with degrees of freedom of $v = 8/\gamma^2$, a mean of $\mu = v$, and a variance of $\sigma^2 = 2v$, where γ is the skewness of the Pearson Type III curve. Therefore, whenever $\gamma > 0$, the p value associated with T is approximated as $\Pr(T) = \Pr(X \leq x)$, where X is a χ^2 random variable with v degrees of freedom and $x = \mu + T\sigma$. In MRPP applications, the skewness is often negative. Therefore, the p value in Equation 8 can be approximated using the complement of the χ^2 CDF, $\Pr(T) = 1 - \Pr(X \leq x)$, where X is a χ^2 random variable with v degrees of freedom and $x = \mu - T\sigma$. When the skewness is zero, the Pearson Type III distribution becomes a standard normal distribution, and the p value of T can be easily obtained from the standard normal CDF.

Commensuration. When the number of dimensions (r) in the outcome observations is greater than one, the de-

pendent variables may be expressed in completely different units, so a multivariate standardization problem arises. Mielke and Berry (1994) proposed two methods of dealing with the commensuration problem. Both methods slightly change the distance function defined in Equation 4. Classical normal-theory-based multivariate procedures tackle this commensuration issue implicitly by formulating test statistics that are invariant under full rank linear transformations of the observed data. They are able to achieve this invariance because all four MANOVA test statistics involve the inverse of the sample variance-covariance matrix so that the resulting test statistics become scale free. In MRPP applications, the commensuration issue becomes explicit, but the goal is still the same as in classical tests—that is, to make the test statistic scale free. It will be shown later that when $v = 2$, Hotelling commensuration leads to a test statistic that is functionally equivalent to the BNP trace for the MANOVA. Therefore, using an explicit commensuration procedure in the MRPP does not change the scientific interpretation of the test result.

The first method is termed *Euclidean commensuration* and is intended for $v = 1$. Returning to the notations used in Equation 1, let $\mathbf{y}_I = (y_{1I}, \dots, y_{rI})'$ denote an $r \times 1$ vector ($r \geq 2$) of noncommensurate multivariate response observations from subject I . The corresponding vector of commensurate multivariate observations $\mathbf{x}_I = (x_{1I}, \dots, x_{rI})'$ can be found by premultiplying the vector \mathbf{y}_I by an $r \times r$ matrix $\Phi^{-1/2}$ —that is, $\mathbf{x}_I = \Phi^{-1/2}\mathbf{y}_I$, where $\Phi = \text{diag}[\phi_j^2]$ denotes a diagonal matrix with elements ϕ_j^2 on the diagonal for $j = 1, \dots, r$. The ϕ_j s are defined as

$$\phi_j = \sum_{I < J} |Y_{jI} - Y_{jJ}|.$$

Combining the results above with Equation 4, the Euclidean commensuration essentially changes the definition of the Euclidean distance function to

$$\Delta_{I,J} = \left[(\mathbf{y}_I - \mathbf{y}_J)' \Phi^{-1} (\mathbf{y}_I - \mathbf{y}_J) \right]^{1/2}. \quad (9)$$

The second method is called *Hotelling commensuration*, to be used with the squared Euclidean distance ($v = 2$). It changes the definition of the distance function to

$$\Delta_{I,J} = (\mathbf{y}_I - \mathbf{y}_J)' \mathbf{S}^{-1} (\mathbf{y}_I - \mathbf{y}_J), \quad (10)$$

where

$$\mathbf{S} = \frac{1}{N} \times \begin{bmatrix} \sum_{I=1}^N (y_{1I} - \bar{y}_1)^2 & \cdots & \sum_{I=1}^N (y_{1I} - \bar{y}_1)(y_{rI} - \bar{y}_r) \\ \vdots & & \vdots \\ \sum_{I=1}^N (y_{rI} - \bar{y}_r)(y_{1I} - \bar{y}_1) & \cdots & \sum_{I=1}^N (y_{rI} - \bar{y}_r)^2 \end{bmatrix}$$

is the $r \times r$ sample variance-covariance matrix.

Relation to existing statistics. When $r = 1$, $v = 2$, and group weights are $C_i = (n_i - 1)/(N - k)$, the MRPP

statistic δ_{obs} has the following relation to the mean square treatment used in the ANOVA (Mielke & Berry, 1994):

$$\delta_{\text{obs}} = \frac{(N - 1)\mu_\delta - 2(k - 1)MS_{\text{trt}}}{N - k}, \quad (11)$$

where

$$MS_{\text{trt}} = \frac{1}{k - 1} \sum_{i=1}^k n_i (\bar{y}_i - \bar{y}),$$

\bar{y}_i is the mean of the i th group, and \bar{y} is the grand mean for the combined sample. Because μ_δ is fixed for a given data set and MS_{trt} depends only on the means, the statistical inference based on the MRPP is not influenced by variance heterogeneity (Mielke & Berry, 1994).

When $r \geq 2$, $v = 2$, group weights are $C_i = (n_i - 1)/(N - k)$, and Hotelling commensuration is used, the MRPP statistic is related to the BNP trace criterion V in the MANOVA by the following identity

$$\delta_{\text{obs}} = \frac{2(r - V)}{N - k}. \quad (12)$$

In addition, all four MANOVA test statistics are equivalent when the number of groups is two. It is also well known that T^2 is a generalization of the Student's t test. Therefore, these parametric tests are all special cases of the MRPP test, yet the MRPP does not make use of any normal distribution theory to make inferences.

MRPP regression analyses of linear models. The MRPP test, as described in the previous section, is applicable only to one-way designs. To analyze data arising from more complex experimental designs, the MRPP is combined with regression models with appropriately reparameterized design matrices. This procedure is called *MRPP regression analyses of linear models* (Mielke & Berry, 2001).

It is well known that one can approach the F test of an effect in factorial ANOVA models within the ordinary least squares (OLS) regression framework by comparing a reduced model (without the effect being analyzed) with a full model (with the effect). For example, a fixed-effects one-way layout posits a full model of the form

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad (13)$$

where y_{ij} denotes the response observation from the j th subject ($j = 1, \dots, n_i$) in the i th group ($i = 1, \dots, k$), μ is the grand mean, α_i is the i th treatment effect, and e_{ij} is the error term. To analyze the treatment effect, a reduced model is given by

$$y_{ij} = \mu + e_{ij}. \quad (14)$$

To obtain the ANOVA F statistic for the α effect, one would first fit the model in Equation 14 and then add the $k - 1$ variables representing the α effect to the regression equation to fit the full model in Equation 13. The F test of incremental R^2 from the reduced model to the full model is equivalent to the F statistic derived using an ANOVA.

The MRPP regression analyses use the same rationale, except that it replaces the parametric F test of incremental fit by the MRPP test. For the preceding one-way layout,

one would first fit the reduced model, obtain the regression coefficients, and compute the residuals, using the regression equation. Then an MRPP analysis of the residuals provides a statistical test of the treatment effect. More generally, MRPP regression analysis is a three-step procedure. In Step One, one would fit a reduced regression model with all the effects in the regression equation but the one being analyzed. In Step Two, residuals are computed from the reduced model. This is followed by Step Three, an MRPP analysis of the residuals. For example, consider the case in which there is a two-way factorial design and the experimenter is interested in analyzing the interaction effect between the two factors. The full model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}, \quad (15)$$

where the α s and β s are the main effects and $(\alpha\beta)_{ij}$ represents the interaction. The reduced model without the interaction is given by

$$y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}. \quad (16)$$

The MRPP analysis of residuals from fitting the reduced model in Equation 16 is a test of the interaction effect. Note that the MRPP test requires grouping information, but for all experimental designs, grouping information is implicit. For example, for a 2×2 factorial design, there are, altogether, four treatment combinations. All that is needed in the MRPP analysis of residuals is a single grouping variable indicating to which treatment combination a particular residual belongs.

Mielke and Berry (2001) suggested that instead of using residuals from OLS regression, one can more profitably use residuals from least absolute deviation (LAD) regression. Unlike the OLS regression, which estimates the regression coefficients by minimizing squared deviations, LAD regression minimizes absolute values of deviations. Mielke and Berry (2001) showed that MRPP analyses of LAD residuals, using the Euclidean distance function ($v = 1$), leads to a more robust test against outliers.

Description of the Macros

Two SPSS macros are provided. The macros are programmed using SPSS's matrix language. The first macro is a fully functional MRPP routine that uses the Pearson Type III approximation to obtain p values. The second macro is a simple LAD regression program that uses an algorithm tailored after the downhill simplex method (Press, Teukolsky, Vetterling, & Flannery, 1992) to obtain LAD regression coefficients. The downhill simplex takes starting values from an initial OLS regression. Together, the two macros can perform all the tests described in the preceding sections. The macros can be downloaded in electronic form from the following Web site: www.unc.edu/~lcai/programs.html.

As is true for all SPSS macro programs, one only needs to run them once in each SPSS session. As soon as the macros are run, they become part of the SPSS system and can be invoked in the same way as any other SPSS procedures. The macros will stay active until the user quits SPSS. The macros assume that the input data set is con-

tained in the working data file. Only limited error checking is provided, so if the macro prints error messages, discard the output and do not interpret the results. The missing data is deleted list-wise.

Macro MRPP. The syntax of the MRPP macro is as follows:

```
MRPP dist=[1|2] comm=[0|1|2] weight=[1|2]
iv=grp dvs=dep vars.
```

The "dist" option determines the type of distance function to be used in the MRPP analysis, where 1 indicates Euclidean distance ($v = 1$) and 2 indicates squared Euclidean distance ($v = 2$). The default value is 1.

The "comm" option determines the type of multivariate commensuration to be applied on the outcome observations, where 0 means no commensuration, 1 means Euclidean commensuration, and 2 means Hotelling commensuration. The default value is 1. However, this option has no effect when the outcome is univariate, because commensuration is not necessary for $r = 1$.

The "weight" option determines the type of group weights used in computing the MRPP test statistic. When it is set to 1, weights of the form $C_i = n_i/N$ will be used, and when it is set to 2, weights of the form $C_i = (n_i - 1)/(N - k)$ will be used. The default value is 1.

The user must specify the name of the grouping (independent) variable in the required argument, "iv." The dependent variables should be listed in the last argument, "dvs." The names of the dependent variables must be separated by blank spaces.

For example, suppose that the SPSS working data file contains data from a multivariate one-way design, that the dependent variables are called "y1," "y2," and "y3," and that the grouping variable is called "grp." The following syntax

```
MRPP iv=grp dvs=y1 y2 y3.
```

conducts an MRPP analysis using Euclidean distance, Euclidean commensuration, and the first type of group weights. Note that by omitting the options, the default values are used. The following syntax

```
MRPP dist=2 comm=2 weight=2 iv=grp
dvs=y1 y2 y3.
```

conducts an MRPP analysis using squared Euclidean distance, Hotelling commensuration, and the second type of group weights. This combination of options produces a permutation version of the MANOVA.

Macro LADREG. The syntax of the LADREG macro is

```
LADREG ftol=[1E-6] maxit=[200] const=[1]
dv=dv ivs=predictors.
```

The "ftol" option sets the convergence criterion for the iterative algorithm. The default value is 1E-6, in scientific notation.

The "maxit" option sets the maximum number of iterations allowed in the iterative line-fitting algorithm before it quits without reaching the minimum. The default is 200.

The “const” option determines whether the intercept is estimated or not. When it is set to 1, a column of ones is included in the design matrix. When it is set to 0, the intercept is not estimated. The default value is 1.

The required argument, “dv,” takes the name of the dependent variable. The dependent variable must be in the SPSS working data file.

The names of the predictors are entered through the “predictors” argument. The names must be separated by blank spaces.

For instance, if the dependent variable is called “y” in the open data set and the predictors are called “x1” and “x2,” the

```
LADREG dv=y ivs=x1 x2.
```

code regresses “y” on “x1,” “x2,” and the intercept.

Examples

In this section, I will analyze two example data sets, using the SPSS macros provided in the last section. The output will be provided, and the relation of the MRPP analyses to their parametric counterparts will be explored. The reader can download the SPSS syntax file generating the output from the same Web address as that for the macro download page.

MRPP analyses of multivariate outcomes. The data set comes from Mielke et al. (1976). In the original manuscript, it is listed in Table I on p. 1419. The data set contains three outcome variables, called “v1,” “v2,” and “v3.” The grouping variable is called “grp.” There are $k = 4$ groups, and the total N is 21. The design is not balanced, and the group sizes (n_i) are 5, 4, 8, and 4. The

```
mrpp dist=1 comm=0 weight=1 iv=grp dvs=v1
v2 v3.
```

command conducts an MRPP analysis using Euclidean distance and the first type of group weight. No commensuration is requested, since the original analysis in Mielke et al. does not use any kind of commensuration. The command above replicates Mielke et al.’s analysis.

The output from the macro is shown in Figure 1. The first few lines print some information about the options. It is seen that no commensuration is invoked and that the Euclidean distance is used. The “Group Summary” table prints the group size (n_i), average within-group distance (ξ_i), and the group weights (C_i). The MRPP statistic $\delta_{\text{obs}} = \sum_{i=1}^k \xi_i C_i = 3.45$ is located underneath “Tests of Significance” as “Delta.” Then the exact moments of the MRPP null distribution are printed as “Mean” ($\mu_\delta = 8.11$), “Variance” ($\sigma^2 = .31$), and “Skewness” ($\gamma = -1.06$). The standardized T statistic is found to be $(\delta_{\text{obs}} - \mu_\delta)/\sigma = (3.45 - 8.11)/\sqrt{.311} = -8.35$, which gives a p value of 2.52E-6. Assuming that the three outcome measures are commensurate, there is indication of large between-group difference. These results match exactly with the output obtained from the BLOSSOM package (Cade & Richards, 2001).

A slightly different command,

```
mrpp dist=1 comm=1 weight=1 iv=grp
dvs=v1 v2 v3.
```

conducts the same analysis, but with Euclidean commensuration turned on. The output is shown in Figure 2. With a standardized statistic of -8.18 and a p value of 3.34E-6, the macro produced exactly the same results as BLOSSOM (Cade & Richards, 2001).

```
Multi-Response Permutation Procedure (MRPP)

No commensuration invoked
Metric Euclidean distance function invoked
Total sample size
      21
Group summary
      N      Distance      Weights
5.000000000000  3.580166225429  .238095238095
4.000000000000  3.214880033153  .190476190476
8.000000000000  3.681188834369  .380952380952
4.000000000000  3.060205742503  .190476190476
Moments of the MRPP null distribution
      Mean      Variance      Skewness
8.110161833941  .311462970802  -1.05618948865
Test of Significance
      Delta      T      Sig.
3.450032614510  -8.35015906239  .000002517390
```

Figure 1. MRPP analysis without commensuration.

```

Multi-Response Permutation Procedure (MRPP)

Euclidean commensuration invoked: print the Phi matrix
  820836      0      0
    0  698896      0
    0      0 1102500

Metric Euclidean distance function invoked

Total sample size
      21

Group summary
      N      Distance      Weights
5.000000000000 .003893040759 .238095238095
4.000000000000 .003491150118 .190476190476
8.000000000000 .004050703860 .380952380952
4.000000000000 .003412210188 .190476190476

Moments of the MRPP null distribution
      Mean      Variance      Skewness
.008679002569 .000000357445 -1.05825645899

Test of Significance
      Delta      T      Sig.
.003784965519 -8.18583042093 .000003343974

```

Figure 2. MRPP analysis with Euclidean commensuration.

```

Multi-Response Permutation Procedure (MRPP)

Hotelling Commensuration invoked: print SSCP matrix
  270.5714286  225.5714286  277.8571429
  225.5714286  233.2380952  264.1904762
  277.8571429  264.1904762  370.9523810

Squared Euclidean distance function invoked

Total sample size
      21

Group summary
      N      Distance      Weights
5.000000000000 .130895548347 .235294117647
4.000000000000 .242633791183 .176470588235
8.000000000000 .289113143558 .411764705882
4.000000000000 .134329829458 .176470588235

Moments of the MRPP null distribution
      Mean      Variance      Skewness
.300000000000 .000441890947 -.515294791321

Test of Significance
      Delta      T      Sig.
.216368532954 -3.97843405036 .000729406618

```

Figure 3. MRPP analysis with Hotelling commensuration.

The command

```
mrpp dist=2 comm=2 weight=2 iv=grp
dvs=v1 v2 v3.
```

uses the squared Euclidean distance with Hotelling commensuration and the second type of group weights for the same one-way design. The output is shown in Figure 3. Again, the results match exactly with the BLOSSOM package (Cade & Richards, 2001). As was noted above, this command essentially produces a permutation version of the MANOVA. In this case, the BNP trace criterion V for the test of group difference is equal to 1.161. The direct correspondence between the MRPP test statistic and V is evident, because according to Equation 12, δ_{obs} should be equal to

$$\frac{2(r - V)}{N - k} = \frac{2(3 - 1.161)}{21 - 4} = .216,$$

which is indeed the case. For this data set, the parametric test V translates into an approximate F statistic [$F(9,51) = 3.577, p = .002$]. Judging from the rather small p values obtained from both the parametric test and the permutation test, it seems that the groups do differ significantly.

MRPP regression analysis of LAD residuals. In this section, I will analyze a data set from the popular experimental design textbook by Keppel (1991). This data set characterizes a 2×3 balanced factorial design. Factor A has two levels, and Factor B has three levels. The cell size is 4, so the total N is 24, and the number of treatment combinations is 6. In this example, I will analyze the interaction effect, using both the LADREG macro and the MRPP macro.

First, the full and reduced models are already given by Equations 15 and 16. For a 2×3 design, three columns of effects code are needed to fully represent the two main effects. In the data set, the outcome variable is called “y;” the effects coded variable for Factor A is called “x1,” and the two coded variables to represent Factor B are called “x2” and “x3.” In addition, I create a six-level grouping variable

(called “tc”) to indicate to which treatment combination the outcome observation belongs. In the next step, “x1,” “x2,” and “x3” are entered into the LAD regression. This is the reduced model, and LAD residuals are computed from the reduced model coefficients. In the final step, I run the MRPP macro, with the residuals as the outcome variable and “tc” as the grouping variable.

The command

```
LADREG dv=y ivs=x1 x2 x3.
```

fits the reduced model to the data. The output is shown in Figure 4. It is seen that the OLS solution is also the LAD solution, because the first row of the Starting Simplex is the same the first row in the Final Simplex. The LAD regression coefficients for the intercept, “x1,” “x2,” and “x3” are 10, -1 , -3 , and 1, in that order. It took the macro five iterations to determine that it had reached a minimum, and at this point, the sum of absolute values of the residuals was 94.⁴ To compute the residuals using the LAD regression coefficient estimates, the following SPSS commands can be used:

```
compute res=y-(10-x1-3*x2+x3).
```

For the MRPP analysis of residuals, the

```
mrpp iv=tc dvs=res.
```

command is used; the output is shown in Figure 5. The MRPP p value for the $A \times B$ interaction is .46. In contrast, the conventional ANOVA of this interaction effect gives $F(2,18) = 3.92, p = .038$, which is significant at the .05 level.

Discussion

The MRPP is an alternative to the traditional normal-theory-based parametric tests. However, MRPP is not widely known to psychological researchers, and part of the reason is that it has not been incorporated into major statistical packages. To overcome this limitation, SPSS macros that implement the MRPP test were provided in

```
Least Absolute Deviation (LAD) Regression Using
Downhill Simplex Minimization Method

Criterion variable is:
Y
Sample Size
24
Number of iterations:
5.00
Sum of absolute values of the residuals:
94.00

LAD Parameter Estimates
Intercep      X1          X2          X3
10.00000    -1.00000    -3.00000    1.000000
```

Figure 4. LADREG output.

Multi-Response Permutation Procedure (MRPP)

```

Commensuration not needed: dimensionality of response = 1
Metric Euclidean distance function invoked
Total sample size
      24
Group summary
      N          Distance          Weights
4.000000000000  4.000000000000  .166666666667
4.000000000000  5.000000000000  .166666666667
4.000000000000  6.000000000000  .166666666667
4.000000000000  7.000000000000  .166666666667
4.000000000000  5.000000000000  .166666666667
4.000000000000  5.000000000000  .166666666667
Moments of the MRPP null distribution
      Mean          Variance          Skewness
5.318840579710  .275221989148  -.754307128559
Test of Significance
      Delta          T          Sig.
5.333333333333  .027625448948  .460765917774

```

Figure 5. MRPP analysis of LAD residuals.

this article, in the hope that they would make MRPP more accessible to psychologists analyzing experimental data. It was shown, by analyzing example data sets, that the macros are as easy to use as the standard SPSS procedures, such as GLM or ONEWAY. The macros, example data sets, and analyses can be downloaded from the following Web site: www.unc.edu/~lcai/programs.html.

REFERENCES

- BOX, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics*, **16**, 769-771.
- BRADLEY, J. V. (1977). A common situation conducive to bizarre distribution shapes. *American Statistician*, **31**, 147-150.
- CADE, B. S., & RICHARDS, J. D. (2001). *User manual for BLOSSOM statistical software*. Fort Collins, CO: Midcontinent Ecological Science Center, U.S. Geological Survey.
- EDGINGTON, E. S. (1966). Statistical inference and nonrandom samples. *Psychological Bulletin*, **66**, 485-487.
- EDGINGTON, E. S. (1995). *Randomization tests* (3rd ed.). New York: Dekker.
- FISHER, R. A. (1935). *The design of experiments*. Edinburgh: Oliver & Boyd.
- GOOD, P. (1993). *Permutation tests: A practical guide to resampling methods for testing hypotheses*. New York: Springer.
- HARTER, H. L. (1967). A new table of percentage points of the Pearson Type III distribution. *Technometrics*, **11**, 177-187.
- HAYES, A. F. (1996). Permutation test is not distribution-free: Testing $H_0: \rho = 0$. *Psychological Methods*, **1**, 184-198.
- JOHNSON, N. L., KOTZ, S., & BALAKRISHNAN, N. (1994). *Continuous univariate distributions* (Vol. I, 2nd ed.). New York: Wiley.
- KEPPEL, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- MACCALLUM, R. C. (1988). Multidimensional scaling. In J. R. Nesselrode & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 421-445). New York: Plenum.
- MICCERI, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, **105**, 156-166.
- MIELKE, P. W. (1984). Meteorological applications of permutation techniques based on distance functions. In P. R. Krishnaiah & P. K. Sen (Eds.), *Handbook of statistics: Vol. 4. Nonparametric methods* (pp. 813-830). Amsterdam: North-Holland.
- MIELKE, P. W., & BERRY, K. J. (1994). Permutation tests for common locations among samples with unequal variances. *Journal of Educational & Behavioral Statistics*, **19**, 217-236.
- MIELKE, P. W., & BERRY, K. J. (2001). *Permutation methods: A distance function approach*. New York: Springer.
- MIELKE, P. W., BERRY, K. J., & JOHNSON, E. S. (1976). Multi-response permutation procedures for *a priori* classifications. *Communications in Statistics: Theory & Methods*, **5**, 1409-1424.
- PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T., & FLANNERY, B. P. (1992). *Numerical recipes in FORTRAN: The art of scientific computing* (2nd ed.). New York: Cambridge University Press.
- WILCOX, R. R. (1997). *Introduction to robust estimation and hypothesis testing*. San Diego: Academic Press.

NOTES

1. A free stand-alone program (BLOSSOM; Cade & Richards, 2001) is available from www.fort.usgs.gov/products/software/blossom/blossom.asp. However, users still face data management problems in transporting their data sets into formats amenable to analysis in BLOSSOM.
2. The notations are primarily formatted after Mielke and Berry (2001), to achieve consistency.
3. For notational simplicity, both the "excess group" and the truncation of distance to a preset maximum are excluded from discussion in the present article (for details, see Mielke & Berry, 2001).
4. Note that the LAD coefficients are not unique.