Mello Moraes 1721, Bl.A, Cid. Universitaria, São Paulo, SP, CEP 05508-900, Brazil). Any comments or suggestions on future improvements are welcome.

Eduardo B. Ottoni
University of São Paulo, Brazil

(Manuscript received July 17, 1995;
revision accepted for publication August 20, 1995.)

PERMUSTAT Randomization Tests for the Macintosh

Most behavioral scientists test hypotheses by using parametric tests designed to determine whether a parameter is different from a value proposed under the null hypothesis. But there is reason to believe that such parametric tests may not be appropriate for many behavioral science applications. Samples are rarely collected randomly from the population, and thus the parametric model, which requires random sampling, simply does not apply. In addition, since parametric tests require random sampling to produce an accurate p value, the computed probabilities may not even be correct (cf. Edgington, 1995). Also, generalization to a population often is not the goal of scientific inquiry. As has been eloquently articulated elsewhere (Hunter & May, 1993; May & Hunter, 1993; Mook, 1983), often scientists are interested not in whether two variables are related in a population but rather in the processes that produce a relationship between these variables in a sample. Since theories that propose causal mechanisms can be tested without reference to a population, why base statistical inferences on the outcome of methods designed to test the value of a parameter? Finally, parametric tests require additional assumptions that sometimes are not met in real behavioral science data (e.g., Bradley, 1977; Micceri, 1989). While some assumption violations do not always invalidate the result of a parametric test, some "real-world" data configurations produce problems, such as heteroscedasticity (Boneau, 1960; Edgell & Noon, 1984; Hayes, 1996; Kowalski, 1973) and nonindependence of sample observations (e.g., time series data or when all data are derived from the same subject; Hayes, 1996; Zimmerman, Williams, & Zumbo, 1992).

Originally introduced by Fisher (1935), the randomization test is ideally suited to testing hypotheses common in behavioral science research. It does not require random sampling from a population and makes no burdensome assumptions such as homoscedasticity or normally distributed errors as long as inference about a parameter is not the goal. The randomization test constructs a p value for a statistic of interest such as the difference between two means by computing the statistic in the obtained data and then repeatedly scrambling the

data set and recomputing the statistic of interest. The p value is the proportion of statistics in the permuted data that equal or exceed the original statistic.

The randomization test is well known to statisticians but has not received much attention outside of statistics literature until recently. Now, the method is being recommended more and more frequently in the behavioral science literature (e.g., Blair & Karniski, 1993; May & Hunter, 1993; Wampold & Worsham, 1986), several books devoted to the topic are available (Edgington, 1995; Good, 1994; Manly, 1991), and discussions of the method are also now beginning to appear in introductory behavioral statistics texts (e.g., May, Masson, & Hunter, 1990).

The convenient aspect of the randomization test is that since p values are computed conditionally on the observed data, the test requires no assumptions about the populations from which the data are sampled. The only assumption of this test is that the dependent or response variable is not systematically associated with the predictor or explanatory variable owing to nonrandom assignment or other factors introduced by the experimenter (see, e.g., Hayes, 1996, for a discussion of the *exchangeability* assumption). The null hypothesis tested is that the obtained relationship between the dependent variable and the independent variable is merely the result of random association. Rejecting this null hypothesis means that there is reason to believe that the obtained association is not just the result of randomness. Although this is a nonparametric conclusion, it is not trivial. Both basic and applied research revolve around determining whether changes in one variable, be they changes in the environment, experimental situation, information subjects receive, and so forth, are associated systematically with changes in some outcome or response. If a researcher can conclude that an obtained association is not just chance association, speculation as to why can begin. Similarly, theory can be tested by determining whether changes in the relevant variables affect the relevant responses as predicted by theory. But, if inference about a parameter is the goal, the randomization test requires the same assumptions of homoscedasticity and independence of observations as traditional parametric tests (Boik, 1987; Hayes, 1996; Keller-McNulty & Higgins, 1987; Rasmussen, 1989).

Until recently, randomization tests have been unavailable to behavioral scientists because most statistical packages do not include randomization routines. However, a few independent software designers have produced programs that can permute and compute very quickly and are available at little cost (e.g., NPStat: May, Hunter, & Masson, 1993). Additionally, several books and journal publications include program code for various randomization tests (e.g., Chen & Dunlap, 1993; Edgington, 1995; Green, 1977; Manly, 1991; Onghena & May, 1995; Siegel & Castellan, 1988).

Unfortunately, most of the available ready-to-run computer software operates only on DOS-based machines, keeping randomization tests out of convenient

reach of scientists who prefer the Macintosh environment. The remainder of this announcement describes PERMUSTAT, a stand-alone application for the Macintosh that will conduct randomization tests for three hypotheses commonly tested in the behavioral sciences.

PERMUSTAT estimates p values for the following tests: (1) the t test that the difference between two means collected from independent samples is zero, (2) the t test that the difference between two means collected from two dependent (repeated measures) samples is zero, and (3) the t test that a Pearson product-moment correlation coefficient is significantly different from zero.

System requirements. PERMUSTAT runs on any Macintosh operating under operating system 6.0.7 or later. The program takes only 105K of disk space and requires no special memory enhancements, graphics capabilities, or math coprocessors. At least 500K of RAM is recommended.

Data entry. Data can be entered into PERMUSTAT manually through the keyboard or by reading an ASCII disk file generated with a word processor, statistical program, or spreadsheet. Data files can be saved to disk.

Method of analysis. PERMUSTAT can compute both approximate and exact p values. For approximate p values, PERMUSTAT uses the logic of the "approximate" randomization test (Edgington, 1995). That is, a random sample of all possible data permutations is used to generate the null frequency distribution of outcomes. The user can specify the number of times the data should be permuted. PERMUSTAT can also compute exact p values by enumerating all possible permutations of the data set appropriate to the test selected.

Output. After data randomization, the computer displays the computed test statistic, the group or variable means (if relevant), and estimates of or the exact one- and two-tailed p values of the statistic. Output can be saved to disk as an ASCII file.

Limitations. PERMUSTAT cannot handle data sets smaller than 3 cases or larger than 500 cases. The program allows little editing of data during manual entry and no editing capabilities are provided after the data are entered.

Speed. PERMUSTAT's speed has been tested on four Macintosh computers with no special mathematical coprocessors. The number of seconds it took each com-

puter to complete the computations, per 1,000 permutations, is displayed in Table 1.

Availability. A current version of PERMUSTAT (Version 1.2) can be obtained through anonymous FTP to the MacPsych archive at ftp.stolaf.edu (see Huff & Sobloff, 1993). Alternatively, a copy of the program and accompanying documentation can be obtained by sending a request to Andrew F. Hayes, Department of Psychology, University of New England, Armidale, NSW 2351, Australia. Within Australia, please include a 3.5-in. diskette and a self-addressed envelope with sufficient return postage.

REFERENCES

- BLAIR, R. C., & KARNISKI, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, **30**, 518-524.
- BOIK, R. J. (1987). The Fisher-Pitman permutation test: A non-robust alternative to the normal theory F test when variances are heterogeneous. *British Journal of Mathematical & Statistical Psychology*, **40**, 26-42.
- BONEAU, C. A. (1960). The effects of violations of assumptions underlying the t test. *Psychological Bulletin*, **57**, 49-64.
- BRADLEY, J. V. (1977). A common situation conducive to bizarre distribution shapes. *American Statistician*, **31**, 147-150.
- CHEN, R. S., & DUNLAP, W. P. (1993). SAS procedures for approximate randomization tests. *Behavior Research Methods, Instruments, & Computers*, **25**, 406-409.
- EDGEELL, S. E., & NOON, S. M. (1984). Effects of violation of normality on the t test of the correlation coefficient. *Psychological Bulletin*, **95**, 576-583.
- EDGINGTON, E. S. (1995). *Randomization tests* (3rd ed.). New York: Dekker.
- FISHER, R. A. (1935). *The design of experiments*. Edinburgh and London: Oliver & Boyd.
- GOOD, P. (1994). *Permutation tests: A practical guide to resampling methods for testing hypotheses*. New York: Springer-Verlag.
- GREEN, B. F. (1977). A practical interactive program for randomization tests of location. *American Statistician*, **31**, 37-39.
- HAYES, A. F. (1996). The permutation test is not distribution-free. Testing $H_0: \rho = 0$. *Psychological Methods*, **1**, 184-198.
- HUFF, C., & SOBILOFF, B. (1993). MacPsych: An electronic discussion list and archive for psychology concerning the Macintosh computer. *Behavior Research Methods, Instruments, & Computers*, **25**, 60-64.
- HUNTER, M. A., & MAY, R. B. (1993). Some myths concerning parametric and nonparametric tests. *Canadian Psychology*, **34**, 384-389.
- KELLER-MCNULTY, S., & HIGGINS, J. J. (1987). Effects of tail weight and outliers on power and Type I errors of robust permutation tests for location. *Communication in Statistics: Simulation & Computation*, **16**, 17-35.
- KOWALSKI, C. J. (1973). On the effects of nonnormality on the distribution of the sample product-moment correlation coefficient. *Applied Statistics*, **21**, 1-12.
- MANLY, B. F. J. (1991). *Randomization and Monte Carlo methods in biology*. London: Chapman & Hall.
- MAY, R. B., & HUNTER, M. A. (1993). Some advantages of permutation tests. *Canadian Psychology*, **34**, 401-407.
- MAY, R. B., HUNTER, M. A., & MASSON, M. E. J. (1993). *NPStat* (Version 3.8) [Computer program]. University of Victoria, Department of Psychology.
- MAY, R. B., MASSON, M. E. J., & HUNTER, M. A. (1990). *Application of statistics in behavioral research*. New York: Harper & Row.
- MICCERI, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, **105**, 156-166.
- MOOK, D. (1983). In defense of external invalidity. *American Psychologist*, **38**, 379-387.

Table 1
Computational Speed of PERMUSTAT Operating on Four Different Computers

Computer	Clock/Processor	Bivariate Sample Size		
		10	30	50
Macintosh SE	16 MHz/68000	82	195	307
Macintosh LC	25 MHz/68020	72	118	165
Macintosh LCIII	25 MHz/68030	29	48	66
Macintosh Centris 610	25 MHz/68LC040	12	20	28

Note—Table entries are in seconds/1,000 permutations for computation of approximate p values.

- ONGHENA, P., & MAY, R. B. (1995). Pitfalls in computing and interpreting randomization test p values: A commentary on Chen and Dunlap. *Behavior Research Methods, Instruments, & Computers*, **27**, 408-411.
- RASMUSSEN, J. L. (1989). Computer intensive correlational analysis: Bootstrap and approximate randomization techniques. *British Journal of Mathematical & Statistical Psychology*, **42**, 103-111.
- SIEGEL, S., & CASTELLAN, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.
- WAMPOLD, B. E., & WORSHAM, N. L. (1986). Randomization tests for multiple baseline designs. *Behavioral Assessment*, **8**, 135-143.
- ZIMMERMAN, D. W., WILLIAMS, R. H., & ZUMBO, B. D. (1992). Correction of the Student t statistic for nonindependence of sample observations. *Perceptual & Motor Skills*, **75**, 1011-1020.

Andrew F. Hayes
University of New England
Biddeford, Maine

(Manuscript received January 9, 1995;
 revision accepted for publication July 17, 1995.)