PROGRAM ABSTRACTS/ALGORITHMS

A program for the generation of data sets possessing desired degrees of score stability

DAVID MARSHALL, H. D. DAY, and JOHN CHRISTY
Texas Woman's University, Denton, Texas 76204

One of the more persistent issues in personality theory concerns the existence of stable behavioral dispositions. While intuition suggests considerable cross-situational consistency in behavior, actual correlations among cross-situational measurements of "trait-dependent" behaviors are perplexingly low (Mischel, 1969). Behavioral stability has most often been quantified by the computation of simple correlation coefficients for multiple assessments of the construct; recently, alternative methods for measurement of stability have been proposed (Epstein, 1979; Day, Marshall, Hamilton, & Christy, Note 1). A difficulty with the evaluation of the construct validity of these new measurement approaches is that there is no consensus regarding the identification of manifestations of traits as opposed to behavioral states (see Allen & Potkey, 1981). A possible solution to this validity problem is to examine the characteristics of the stability statistics when applied to computer-generated data sets containing a degree of stability that is mathematically defined.

This paper describes a program that is useful in the study of statistics purporting to measure behavioral stability (Day, Marshall, Hamilton, & Christy, in press, Note 1). The program generates data sets analogous to the setting in which a construct is repeatedly measured on a sample of subjects, and stability is produced by inducing a desired correlation (ρ) between all adjacent pairs of trials in the resulting case by trials (n by k) matrix. In addition, the scores for the n cases within each of the k trials are normally distributed z scores.

Each n by k data set is generated by a two-step process (Knuth, 1981). Initially, independent normal deviates are derived by the polar method. At Step 1, a given data set contains k columns of independent (uncorrelated) z scores for n cases. Step 2 involves the induction of a desired degree of correlation among the k columns. An iterative transformation of the independent z scores is employed, of the form $Y_1 = b_1 + a_{11}X_1$, $Y_2 = b_2 + a_{21}X_1 + a_{22}X_2$, . . . , $Y_n = b_n + a_{n1}X_1 + a_{n2}X_2 + \dots + a_{nn}X_n$, where $b_i = \mu = 0$, X_i s are independent normal deviates, and Y_i s are normal deviates with a given covariance matrix (c_{ij}).

David Marshall and John Christy are with the Department of Mathematics, Computer Science, and Physics; H. D. Day is with the Department of Psychology and Philosophy.

Using $c_{ij} = 1$ if $i = j$, $c_{ij} = \rho$ if $i \neq j$, where ρ is a given correlation constant, and requiring $[(a_{ij})(\text{transpose of } (a_{ij}))] = (c_{ij})$, where (a_{ij}) is the lower triangular matrix of coefficients of the X_i given above, the following results are obtained:

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ii}X_i = \sum_{k=1}^i a_{ik}X_k,$$

where the a s are given by

$$a_{ii} = \sqrt{(1 - \rho)[1 + (i - 1)\rho] / [1 + (i - 2)\rho]}$$

for $i = 1, 2, 3, \dots, n$ and

$$a_{ik} = \rho \sqrt{(1 - \rho) / [1 + (k - 1)\rho] [1 + (k - 2)\rho]}$$

for $i = 2, 3, 4, \dots, n$ and $0 < k < i$ and ρ is a given correlation constant.

Thus, a given dependent z score is obtained as a weighted sum of all independent z scores up to (to the left of) and including the corresponding z score in the k column matrix of independent z scores. The first dependent z is given as simply the first independent z , as its weight reduces to 1. The second dependent z is given as a_{21} times X_1 plus a_{22} times X_2 , and so on. The final result is a collection of scores (trials) that are sampled from normal populations having certain population covariances.

We have tested the efficacy of this algorithm by examining the distributions of trials and the intertrial correlations in numerous generated data sets, each containing scores for 200 cases across 100 trials. Ten data sets each were generated for ρ values varying from 0 to .09 in increments of .01 and for ρ values of .10 to .90 in increments of .10. The following statistical analyses were performed within each data set: Kolmogorov-Smirnov one-sample distribution test on each trial, to assess assumed normality of trials, and comparison of all possible intertrial correlations with the 95% confidence interval on ρ . Regardless of the value of ρ , the following results were consistently obtained: (1) Ninety-six percent to 100% of the trials in a set were normally distributed; (2) 96% to 100% of the serial intertrial correlations ($r_{12}, r_{23}, \dots, r_{k-1,k}$) varied around the desired ρ value within the 95% confidence interval on ρ ; and (3) 95% to 99% of all other intertrial correlations [$r_{ij}, j > (i + 1)$] varied around ρ within the 95% confidence interval on ρ . Sixty percent of all data sets contained 100% each of normal trials and serial intertrial correlations within the 95% range of ρ . When irregularities did occur, the percentages of trials failing

to be normally distributed or of intertrial correlations outside the 95% range of ρ were small and within expected, acceptable limits. The modal percentage of trials failing to be normally distributed was 2%. Modal percentages of serial and nonserial intertrial correlations falling out of range were 1% and 3%, respectively. No significant tendencies for the several types of irregularities to cooccur systematically were detected. Pearson correlations were computed (for data sets containing any irregularities) between the total number of non-normal trials and the total numbers of serial and nonserial intertrial correlations falling out of range, and between the total numbers of serial and nonserial out-of-range correlations. The three resulting Pearson coefficients were .08, .09, and .03, respectively. Based on the foregoing analyses, we conclude that the algorithm produces data sets of sufficiently high quality for use in simulation studies in which normally distributed repeated-measurement data containing known degrees of score stability are required.

Computer System. The program is written in FORTRAN for the DECsystem-2050 and is adaptable to virtually any system. Input consists of the specification values for n , k , ρ , and the number of desired data sets. Each n by k matrix is written into a file for subsequent analysis. Two versions of the program are available. One version is compiled by FORTRAN-20 on the DEC-20 and is interactive, taking parameters from the user at a terminal as the program runs. The other version is generally adaptable to a batch-oriented FORTRAN

environment, and input parameters are expected from cards or other media. Both versions are documented with comments to assist the user in understanding the algorithm and in making necessary changes to the code.

Availability. Source code listings are available at no cost from David Marshall, Department of Mathematics, Computer Science, and Physics, Texas Woman's University, Denton, Texas 76204.

REFERENCE NOTE

1. Day, H. D., Marshall, D., Hamilton, B., & Christy, J. *A comparison of approaches to the measurement of behavioral stability*. Paper presented at the 89th annual meeting of the American Psychological Association, Los Angeles, California, August 1981.

REFERENCES

- ALLEN, B. P., & POTKAY, C. R. On the arbitrary distinction between states and traits. *Journal of Personality and Social Psychology*, 1981, 41, 916-928.
- DAY, H. D., MARSHALL, D., HAMILTON, B., & CHRISTY, J. Some cautionary notes regarding the use of aggregated scores as a measure of behavioral stability. *Journal of Research in Personality*, in press.
- EPSTEIN, S. The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, 1979, 37, 1097-1126.
- KNUTH, D. E. *The art of computer programming* (Vol. 2). *Seminumerical algorithms*. Reading, Mass: Addison-Wesley, 1981.
- MISCHEL, W. Continuity and change in personality. *American Psychologist*, 1969, 24, 1012-1018.

(Accepted for publication July 29, 1982.)