A QUICKBASIC PROGRAM FOR GENERATING CORRELATED MULTIVARIATE RANDOM NORMAL SCORES

HERMAN AGUINIS
University of Colorado at Denver

A QuickBASIC program for generating multivariate random normal scores with given intercorrelations is described. The program uses the Box and Muller and Choleski Factorization algorithms and runs on IBM and IBM-compatible personal computers. A diskette containing the compiled run time and source code versions of the program is available at no charge from the author.

This article describes a program for generating k random vectors **X** of order n, and length N from a multivariate normal population with any specified correlation matrix, **R**. The program was written in QuickBASIC, release 4.5, on a 386 SX microcomputer, and runs on IBM and IBM-compatible personal computers (PCs).

The program can be employed by researchers using Monte Carlo techniques. Until now, researchers interested in generating multivariate correlated scores had the limitations to either use a mainframe computer (International Mathematical and Statistics Library, 1982, subroutine GGNSM) or limit the number of generated variables to $n = 2$ (Alliger, 1992). The present program overcomes these limitations and allows researchers to generate multidimensional vectors of correlated normal scores on a PC. There are several possible applications for this program. For example, repeated samples from a population with a specified correlation matrix can be generated, and these data can then be manipulated (e.g., range restricted on one or more variables). Then the impact of the manipulation can be assessed. For this particular example,

I gratefully acknowledge Gerardo D. Aguinis for his assistance in the algebraic manipulations at the initial stages of this project. Also, I would like to thank Michael A. Fontana (president of the University of Colorado at Denver student chapter of ACM) for his assistance in QuickBASIC programming.

Educational and Psychological Measurement, Vol. 54 No. 3, Fall 1994 687-689

the distribution of *r*s from range-restricted samples could be compared with the distribution of *r*s from nonrestricted samples (e.g., Millsap, 1989).

The Program

The user is prompted interactively to input the desired sample size N, the values for the elements above the diagonal in the population correlation matrix **R** of order n (i.e., $\rho_{1,2}, \rho_{1,3}, \rho_{2,3}, \ldots, \rho_{n-1,n}$), and the desired number of vectors k (i.e., number of samples for each variable). The program allows for an initial $N \leq 1{,}000$ and $n \leq 10$, but these features are easily modifiable. The value for the number of samples k is only limited by the available storage space for the generated data.

Each vector **X** is generated using the Choleski Factorization Method (Moonan, 1957), which has been found to be simpler (Scheuer & Stoller, 1962) and requires less execution time and memory space (Barr & Slezak, 1972) than other multivariate normal generation algorithms. The program functions by first reducing a specified symmetric correlation matrix **R** into the product of a lower triangular matrix **C** and its transpose, $\mathbf{R} = \mathbf{C}'\mathbf{C}$. Then, a random normal vector $\mathbf{Y} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$ is generated using the Box and Muller (1958) algorithms. Finally, a vector **X** is obtained by multiplying the generated vector **Y** by the lower triangular matrix **C** (i.e., $\mathbf{X} = \mathbf{YC} \sim \mathrm{N}(\mathbf{0}, \mathbf{R})$). The computation of **C** for a given **R** is achieved via the square root method (Graybill, 1969; Pearson, 1948).

Initial accuracy tests showed that the observed sample correlations corresponded to the specified parameters. For example, with samples ($k = 1{,}000$) of $N = 100$, $n = 3$, $\rho_{12} = .40$, $\rho_{13} = .60$, and $\rho_{23} = .00$, the observed mean correlations were $\bar{r}_{12} = .398$, $\bar{r}_{13} = .599$, and $\bar{r}_{23} = .002$. Additional tests of sampling adequacy were conducted by examining the variability of the observed correlations. For the same example, the standard deviations of the observed correlations were $SD_{12} = .085$, $SD_{13} = .064$, and $SD_{23} = .103$, which correspond to expected *SD*s of .084, .064, and .101, respectively.

Program Output

Output from the program was designed in a similar fashion to Alliger (1992). An output file (DATA.OUT) includes the correlated multivariate random normal scores. A second output file (STATS.OUT) includes the observed means, observed standard deviations, and observed intercorrelations for each vector **X**.

Program Availability

The executable (MULTIVAR.EXE) and the source (MULTIVAR.BAS) versions of the program are available at no cost on either a 3.5-inch or a

5.25-inch diskette (double or high density). Users who want to obtain the program should send a blank formatted diskette and a self-addressed, stamped envelope to Herman Aguinis, Ph.D., Department of Psychology, University of Colorado at Denver, Campus Box 173, P.O. Box 173364, Denver, CO 80217-3364.

References

Alliger, G. M. (1992). Generating correlated bivariate random normal standard scores in QuickBASIC. *Educational and Psychological Measurement, 52*, 107-108.

Barr, D. R., & Slezak, N. L. (1972). A comparison of multivariate normal generators. *Communications of the ACM, 15*, 1048-1049.

Box, G.E.P., & Muller, M. E. (1958). A note on the generation of random normal deviates. *Annals of Mathematical Statistics, 29*, 610-613.

Graybill, F. A. (1969). *Introduction to matrices with applications in statistics.* Belmont, CA: Wadsworth.

International Mathematical and Statistical Libraries. (1982). *IMSL library reference manual* (Vol. 2, 9th ed.). Houston, TX: Author.

Millsap, R. E. (1989). Sampling variance in the correlation coefficient under range restriction: A Monte Carlo study. *Journal of Applied Psychology, 74*, 456-461.

Moonan, W. J. (1957). Linear transformation to a set of stochastically dependent normal variables. *American Statistical Association Journal, 52*, 247-252.

Pearson, E. S. (Ed.). (1948). *Tracts for computers, No. 25.* Cambridge, England: Cambridge University Press.

Scheuer, E. M., & Stoller, D. S. (1962). On the generation of normal random vectors. *Technometrics, 4*, 278-281.