A QUICKBASIC PROGRAM FOR GENERATING
CORRELATED MULTIVARIATE RANDOM NORMAL SCORES

HERMAN AGUNIS
University of Colorado at Denver

A QuickBASIC program for generating multivariate random normal scores with
given intercorrelations is described. The program runs on DOS and Muller and (Chebyshev)
Factorization techniques 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 tech-
niques. 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 OQNSM) or limit the
number of generated variables to n = 2 (Alligter, 1982). The present program
overcomes these limitations and allows researchers to generate multidimen-
sional vectors of correlated normal scores on a PC. There are several possible
applications for this program. For example, repeated samples from a popu-
lation 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. Apinafor his assistance in the algebraic manipulations
at the initial stages of this project. Also, I would like to thank Michael A. Fentress (president of
the University of Colorado at Denver student chapter of ACME) for his assistance in QUICKBASIC
programming.

Educational and Psychological Measurement, Vol. 54 No. 3, Fall 1994 627-629

687
the distribution of $r$ from range-restricted samples could be compared with
the distribution of $r$ from nonrestricted samples (e.g., Millsap, 1989).

The Program

The user is prompted interactively to input the desired sample size, the
values for the elements above the diagonal is the population correlation
matrix $R$ of order $n^2$ (i.e., $p_1, p_2, p_3, \ldots, p_{n^2}$), and the desired number
of vectors $k$ (i.e., number of samples for each variable). The program allows
for an initial $N = 1,000$ and $n = 10$, 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 Cholesky Factorization Method
(Moosan, 1957), which has been found to be simpler (Scheuer & Stoller,
1962) and requires less execution time and memory space (Barr & Stieglitz,
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, $R = CC'$. Then,
a random normal vector $Y\sim N(0,1)$ is generated using the Box and Muller
(1958) algorithms. Finally, a vector $X$ is obtained by multiplying the gener-
ated vector $Y$ by the lower triangular matrix $C$ (i.e., $X = YC\sim N(0, R)$). 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 corre-
sponded to the specified parameters. For example, with samples ($k = 1,000$)
N = 100, $n = 3$, $p_{12} = .40$, $p_{13} = .60$, and $p_{23} = .00$, the observed mean
correlations were $r_{12} = .398$, $r_{13} = .599$, and $r_{23} = .002$. 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 $SDs$ 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 intercorrela-
tions 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 173369, Denver, CO 80217-3364.

References


