AUTHOR: Thomas, David B.
TITLE: An APL/1500 Procedure for Generating Random Samples from Trivariate Normal Populations.
PUB DATE: 72
NOTE: 14p.
EDRS PRICE: MF-$0.65 HC-$3.29
DESCRIPTORS: *Computer Assisted Instruction; *Computer Programs; Learning; Measurement Techniques; Models; Research; Sampling; *Simulation; *Teachers; *Training

ABSTRACT: The uses of simulation as a technique for instruction and research have been expanding over the past ten years. As an aid to educational practitioners, the sampling procedures associated with trivariate normal populations are described along with a description of an APL/1500 computer program for this level of multivariate sampling in educational research. A sample use of the program is also given. (Author)
AN APL/1500 PROCEDURE FOR GENERATING RANDOM SAMPLES
FROM TRIVARIATE NORMAL POPULATIONS

David B. Thomas

1972
Florida State University
ABSTRACT

The uses of simulation as a technique for instruction and research have been expanding over the past ten years. As an aid to educational practitioners, the sampling procedures associated with trivariate normal populations are described along with a description of an APL/1500 computer program for this level of multivariate sampling in educational research. A sample use of the program is also given.
AN APL/1500 PROCEDURE FOR GENERATING RANDOM SAMPLES
FROM TRIVARIATE NORMAL POPULATIONS

David B. Thomas

Introduction

The uses of simulation as a technique for instruction and research have been expanding over the past ten years. Simulation of psychological theories and the development of "learning machines" have been described by Feigenbaum and Feldman (1963) and others (Green, 1969; Snyder, 1968). Instructional simulations have been suggested as supplements to expository instructional techniques (Raser, 1969). In these instructional applications a formal process from physics, chemistry, or mathematics is modeled and presented to the student as an exercise at a cathode ray tube terminal (Kromhout, Hansen, & Schwarz, 1970). Social science instructional applications are described by Boocock and Schild (1968) and by Raser (1969). In these uses, the social science process may be modeled and presented as a learning game in which the student may act out various roles according to the rules of the game. The use of simulation in teacher training programs is described by Smith and Smith (1966).

At the Florida State University Center for Computer-Assisted Instruction, both the Coursewriter II (IBM, 1968) and APL (Iverson, 1966) programming languages are employed for research in instruction. The
FSU-APL system is implemented on an IBM 1800-based 1500 system which supports both cathode ray tube and teletype-typewriter terminals. The APL/1500 system (Krueger and McMurchie, 1968) permits easy implementation of simulation models based on complex mathematical algorithms. When coupled with the interactive nature of APL, the system provides a simulation capability which may be utilized both for research on instruction and as a means of instruction in mathematics, physics, chemistry, library science, and the social sciences.

An investigation of an innovative teacher training program, with one purpose being the implementation and simulation of a training model on a computer system, is described by King (1970). In this investigation, a methodology was required whereby a normal distribution with a specified mean and variance could be randomly sampled. A method was available which allowed for sampling from these populations. The APL function available employed the method reported by Muller (1959), modified so as to reduce the amount of computer space necessary for execution. Using the function, independent samples could easily be drawn from any normal population with a user specified mean and variance. This APL program is reported by Lippert, et al. (1970) as FSU APL/1500 function FSNU12.

The APL function (FSNU12) was adequate for drawing a series of independent normally distributed random variables. However, as is the case with much research in education and psychology, multiple measurements on the same experimental unit or person were desired. These measurements, rather than being independent, are correlated. In order to simulate phenomena in which multiple measurements on the same persons...
were present, a method was required which would sample from a specified
multivariate normal population. In such a population, persons are
represented as points in an n-dimensional space, wherein each of the
variables is considered a dimension. (A multivariate normal population
was desired because the majority of statistical measures in education
and psychology require the assumption of normality (Morrison, 1967)).

In the multivariate normal population the random variables
are related to one another through the variance-covariance matrix \( \Sigma \).
This matrix, in the case of standardized variables, specifies the inter-
correlations among the variables. By specifying \( \Sigma \), one can then
interrelate the variables (i.e., the measures on persons) as desired.

The simulation of the teacher training model used by King
(1970) required three measures on each person, so that it was necessary
to draw random samples from a trivariate normal population. Tocher
(1963) presented a method for selecting samples from the bivariate
normal population when independently distributed \( N(0,1) \) random variables
were available. The remainder of this paper describes the rationale
for extending Tocher's method, and presents the equations used in the
APL/1500 functions to sample from the trivariate normal populations.
Examples of the function's use are also given.

**Sampling Procedure**

Suppose that the trivariate normal population of interest has
the following variance-covariance matrix, \( \Sigma \), and mean vector \( \mu \)

\[
\Sigma = \begin{bmatrix}
1 & \rho_{12} & \rho_{13} \\
\rho_{21} & 1 & \rho_{23} \\
\rho_{31} & \rho_{32} & 1
\end{bmatrix}
\]

\[
\mu = [0 \ 0 \ 0]
\]
in which the $\rho_{ij}$ are the population values of the correlations between variables $i$ and $j$. If $Y_1$, $Y_2$, and $Y_3$ are three normally and independently distributed random variables, each $N(0, 1)$, then the trivariate normal random variables $X_1, X_2, X_3$ may be represented as functions of the $Y_i$ as follows:

(1) $X_1 = Y_1$
(2) $X_2 = a_1 Y_1 + a_2 Y_2$
(3) $X_3 = a_3 Y_1 + a_4 Y_2 + a_5 Y_3$

The task is to compute the constants $a_i$ such that a specific variance-covariance matrix, $\Sigma$, of the variables $X_1, X_2, X_3$ results.

For any two normally and independently distributed random variables $Y_i$ and $Y_j$, it is known (Anderson, 1958; Hogg and Craig, 1970) that

(4) $E[Y_i Y_j] = 0$.

For correlated standardized random variables $Y_i$ and $Y_j$

(5) $E[Y_i Y_j] = \rho_{ij}$.

From the requirement that the $Y$'s be distributed normally with mean zero and variance 1, it is know that

(6) $E[Y_i^2] = 1$

From these properties, the values of the weights $a_i$ can be derived as follows:

$$\rho_{12} = \frac{E[X_1 X_2]}{\sqrt{E[X_1^2] E[X_2^2]}}$$

from (5)

$$= \frac{E[Y_1(a_1 Y_1 + a_2 Y_2)]}{\sqrt{E[Y_1^2] E[(a_1 Y_1 + a_2 Y_2)^2]}}$$

from (1) and (2)

So that

(7) $\rho_{12} = \frac{a_1}{\sqrt{1}}$

by expansion and from (4) and (6)
Now to derive $a_2$, 

\[
1 = E[X_2^2] 
= E[(a_1Y_1 + a_2Y_2)^2] \quad \text{from (6)}
\]

and in a similar manner, we show:

\[
a_2 = \sqrt{1 - \rho_{12}^2} \quad \text{(8)}
\]

To derive $a_3$, we set:

\[
\rho_{13} = E[X_1X_3] 
= E[Y_1(a_3Y_1 + a_4Y_2 + a_5Y_3)] \quad \text{from (5)}
\]

from which

\[
\rho_{13} = a_3 \quad \text{(9)}
\]

To derive $a_4$, set

\[
\rho_{23} = E[X_2X_3] \quad \text{from (5)}
\]

\[
= E[(a_1Y_1 + a_2Y_2)(a_3Y_1 + a_4Y_2 + a_5Y_3)], \quad \text{from (2) and (3)}
\]

and by procedures similar to those above, we find

\[
\rho_{23} = a_1a_3 + a_2a_4 
= \rho_{12}\rho_{13} + a_4\sqrt{1 - \rho_{12}^2} \quad \text{from (7), (8), and (9)}
\]

from which

\[
a_4 = \frac{\rho_{23} - \rho_{12}\rho_{13}}{\sqrt{1 - \rho_{12}^2}} \quad \text{(10)}
\]

finally, to compute $a_5$:

\[
1 = E[X_3^2] \quad \text{from (6)}
\]

\[
= E[(a_3Y_1 + a_4Y_2 + a_5Y_3)^2] \quad \text{from (3)}
\]
by expanding and substituting we can show that

\[
1 = \rho_{13}^2 + (\rho_{23} - \rho_{12} \rho_{13})^2 + a_5^2
\frac{1}{1 - \rho_{12}^2}
\]

from which

\[
a_5 = \sqrt{1 - \rho_{13}^2 - (\rho_{23} - \rho_{12} \rho_{13})^2}
\frac{1}{1 - \rho_{12}^2}
\]

To summarize,

\[
(7) \quad a_1 = \rho_{12}
\]

\[
(8) \quad a_2 = \sqrt{1 - \rho_{12}^2}
\]

\[
(9) \quad a_3 = \rho_{13}
\]

\[
(10) \quad a_4 = \frac{\rho_{23} - \rho_{12} \rho_{13}}{\sqrt{1 - \rho_{12}^2}}
\]

\[
(11) \quad a_5 = \sqrt{1 - \rho_{13}^2 - (\rho_{23} - \rho_{12} \rho_{13})^2}
\frac{1}{1 - \rho_{12}^2}
\]

From the foregoing, the following restrictions emerge:

(1) \(\rho_{12}^2 = 1\) in order that \(a_4\) may be computed (has finite bounds)

(2) \(\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2 \leq 1 + 2 \rho_{12} \rho_{13} \rho_{23}\) (in order that \(a_5\) remain real).

The \(X_i\) may now be computed from the already selected \(Y_i\). The \(X_i\) will then constitute a random sample from the trivariate normal population with the variance-covariance matrix \(\Sigma\) and mean vector \(\mu\).
The function \( M \) employs a modification of the Box-Muller method for generating standardized normal random deviates (see Muller, 1959). These are then transformed into random variables from a normal distribution with given mean \( \mu \) and variance \( \sigma^2 \) before being output. See Lippert, et al. (1970) for a description of the use of this function independently of the current application (FSNU11). A listing of the function \( M \) follows.

\[
\begin{align*}
\text{V}[N][1] & \quad \text{V} \quad R \leftarrow \text{N} \\ & \quad R \leftarrow (R \ast \text{M}[1]) \ast (R \ast \text{M}[2] \ast 0.5) \ast (1002 \ast (R \ast \text{M}[1]) \ast (R \ast 0.5)) \ast (2 \ast \text{R} \ast ?(N-N, 1 \ast \text{R}) \ast \text{R}) \ast \text{R} \ast 8388608) \ast 0.5
\end{align*}
\]

The function \( X \) computes the constants \( a_i \) and derives the trivariate sample points \( x_1, x_2, x_3 \) by employing the function \( M \) to generate independent \( N(0,1) \) variables which are manipulated according to the procedure given in equations (1)-(3) from the previous section. The explicit result of \( X \) is an \( r \times 3 \) matrix of data in which each row corresponds to the standardized coordinates of the sample point in three dimensional space; \( r \) sample points are generated. Other straightforward measures are required to transform these coordinates into data with given means and variances (i.e., these data are standardized normal variates). A listing of this function follows:

\[
\begin{align*}
\text{V}[X][1] & \quad \text{V} \quad R \leftarrow \text{N} \\ & \quad R \leftarrow (+(C \ast x)/2) \ast 1+2 \times x / C \ast \text{R} \ast 6 \\
[1] & \quad A \leftarrow C[1], B, C[2, D], (1-(C[2] \ast x)+D+(C[3] \ast x/C[\{1,2\}]) \ast B+(1-C[1] \ast 2) \ast 0.5) \ast 2) \ast 0.5
\end{align*}
\]

\[
\begin{align*}
[2] & \quad R[;3] \leftarrow (R \ast N \ast 3 \ast 3 \ast R \ast 0 \ast 1) \ast x \ast 3 \ast 3 \\
[3] & \quad R[;2] \leftarrow R[;1 \ast 2] \ast x \ast 2 \ast A \\
[4] & \quad R \leftarrow 0 \\
[5] & \quad R \leftarrow 0, 0 \ast R \leftarrow '\text{CORRELATIONS OUTSIDE DOMAIN OF FUNCTION}.'
\end{align*}
\]
Example Use

Suppose that a sample of size 20 drawn from a trivariate population was required with the following covariance matrix:

\[
\Sigma = \begin{bmatrix} 1.0 & .5 & .1 \\ .5 & 1.0 & -.6 \\ .1 & -.6 & 1.0 \end{bmatrix}
\]

The following APL expression would generate the required sample:

\[20 \times .5 \times .1 \times -.6\]

Sample data are given below:

<p>| | | | | | | | | | | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7454</td>
<td>2.019</td>
<td>0.1944</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.796</td>
<td>0.8288</td>
<td>0.01032</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.943</td>
<td>0.471</td>
<td>1.363</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4799</td>
<td>0.4117</td>
<td>1.041</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.236</td>
<td>0.441</td>
<td>0.6187</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.23</td>
<td>0.3218</td>
<td>0.3429</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3484</td>
<td>1.037</td>
<td>0.521</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8126</td>
<td>1.79</td>
<td>1.117</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.023</td>
<td>0.004035</td>
<td>0.2626</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7774</td>
<td>1.033</td>
<td>0.02037</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4948</td>
<td>0.01833</td>
<td>0.2669</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.648</td>
<td>1.411</td>
<td>0.4178</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1092</td>
<td>0.2843</td>
<td>0.3142</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1421</td>
<td>0.1251</td>
<td>0.3569</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.118</td>
<td>0.06234</td>
<td>1.663</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9683</td>
<td>0.5068</td>
<td>0.2334</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2228</td>
<td>1.664</td>
<td>2.582</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5176</td>
<td>0.9099</td>
<td>0.4406</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.09321</td>
<td>0.06726</td>
<td>0.464</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7093</td>
<td>1.898</td>
<td>1.597</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The sample correlation matrix for these data is:

\[
R = \begin{bmatrix} 1.00 & .47 & .09 \\ .47 & 1.00 & -.62 \\ .09 & -.62 & 1.00 \end{bmatrix}
\]

These values are reasonable given the size of the sample and the usual sampling errors.
It is hoped that these procedures and their logical extensions will prove useful to behavioral scientists engaged in training research using simulation as a tool. At the present time, educational usages of this nature are limited. These functions were employed in an investigation of a teacher training model (King, 1970). Hopefully, publication of these functions will stimulate additional applications of the technique.
REFERENCES


International Business Machines Corporation. 1500 coursewriter II authors guide. San Jose: IBM Corp., 1968.


