A method to simulate nonnormal distributions has been proposed by Fleishman (1978). This method generates data based on the conceptualization that distributions of variables can be represented by their first four moments, i.e., mean, variance, skewness, and kurtosis. Fleishman's procedure basically transforms normal deviates, X with N(0,1), to a linear function of X up to its cube for univariate nonnormal distribution. More specifically, \( Y = a + bX + cX^2 + dX^3 \). Vale and Maurelli (1983) extended Fleishman's procedure to generate multivariate random numbers with pre-specified intercorrelation structure in addition to the first four moments. To incorporate the desired correlation matrix in simulating multivariate nonnormal distributions with the first four moments within reasonable range, an intermediate correlation matrix needs to be determined. This procedure involves solving a third-degree polynomial. A SAS/IML program, SIMNNM.SAS, has been developed using Vale and Maurelli's procedure to simulate multivariate nonnormal data.

Monte Carlo studies using three variables were conducted to evaluate this procedure in generating nonnormal data. The three variables were assumed to have low, medium, and high correlation, or correlation coefficients of .2, .5, and .8, respectively. Three combinations of skewness and kurtoses were also assumed. Skewness is ranged from 0 to 1 while kurtoses ranged from 0 to 3. Nonnormal data based on certain combination of correlation matrix and nonnormal distribution were then simulated with samples sizes of 200, 500, and 1000. A hundred samples, or replications, were generated for each of the (3 correlation matrices x 3 nonnormal distribution x 3 sample sizes =) 27 combinations. Based on the means of skewness and kurtoses across 100 replications, this procedure generates nonnormal data with smaller skewness and kurtoses than the hypothesized values. The correlation coefficients obtained from the simulated data were, however, quite accurate. The biasness on nonnormal conditions reflected by skewness and kurtoses seemed to be positively related with the size of correlation, especially for kurtoses. The size of nonnormal condition also have impact on the degree of biasness. Sample sizes, however, did not seem to create significant impact on the biasness.

References

START FLEISH(NN,NV,X,R,SKW,KURT);
PRINT "BEGINNING FLEISH ROUTINE. ";
EPS=1.0/(10.0**10);
COE=I(4,NV,0);
DO A = 1 TO NV;
SK=SKW[1,A];
KU=KURT[1,A];
DO E = 1 TO A;
IF (E < A) THEN DO;
  IF (SK < KURT[1,E]) & (KU < KURT[1,E]) THEN DO;
    B = COE[2,E];
    D = COE[4,E];
    GO TO SKDONE;
  END;
END;
B = 1.0;
D = 0.01;
K = 0;
UPDATE:
  K = K + 1;
  IF (K > 50) THEN DO;
    PRINT "FLEISHMAN COEFFICIENTS NOT FOUND.";
    RETURN;
  END;
RUN FUNSK(B,D,SK,1,FSO);
RUN FUNSK(B,D,KU,2,FKO);
IF(ABS(FSO) < EPS & ABS(FKO) < EPS) THEN GO TO SKDONE;
EPB = EPS * B;
EPD = EPS * D;
BU = B + EPB;
DU = D + EPD;
RUN FUNSK(BU,D,SK,1,FSB);
RUN FUNSK(BU,D,KU,2,FKB);
DSB = (FSB-FSO)/EPB;
DSD = (FSB-FSO)/EPD;
RUN FUNSK(BU,D,SK,1,FSB);
RUN FUNSK(BU,D,KU,2,FKB);
DSB = (FSB-FSO)/EPB;
DSD = (FSB-FSO)/EPD;
RC = J(PI,3,999);
QR = Q#F#R + EPS;
QR = (QR > 0);
DO A = 1 TO PI;
  IF (QR[A,1] > 0) THEN DO;
    TH = ARCOS(R[A,1]/SQRT(Q[A,1]*3));
    QT2 = 2*SQRT(Q[A,1]*3);
    RC[A,1] = QT2*CO3(TH/3);
    RC[A,2] = QT2*CO3(TH+2*PI/3);
    RC[A,3] = QT2*CO3(TH+4*PI/3);
  END;
ELSE DO;
  TH = SQRT(R[A,1]*2-Q[A,1]*3)+ABS(R[A,1])*3*(1.0/3);
  SGNR = 1;
  IF R[A,1] < 0 THEN SGNR = -1;
  RC[A,1] = SGNR*(TH+Q[A,1]*3);
END;
END;
END;
RC = RC-AD3*J(I,3,1);
T = ABS(RC);
T = T[A,1] = 2 TO NV;
A1 = A-1;
DO B = 1 TO A1;
K = K + 1;
I = T[K,1];
TT = RC[K,1];
RHO[A,B] = TT;
END;
END;
RHO = RHO + RHO' + I(NV);
U = ROOT(RHO);
X = X*T;
T = I(NN,1,1);
Y = X*T*COE[2,1] + X*X*T*COE[3,1] + X*X*X*T*COE[4,1];
X = X + T*COE[1,1];
PRINT "FLEISHMAN TRANSFORMATION COMPLETED. ";
FINISH;
START READIN(SS,NN,NV,SEED);
PRINT "BEGINNING READIN ROUTINE. ";
NR = NROW(SS);
NC = NCOL(SS);
NV = NC;
PRINT NR NC SS;
NN = NR;
RETURN;
END;
ELSE DO;
\[JJ=I(1, NV, 0);\]
\[MEAN = JJ;\]
\[VAR = JJ;\]
\[SKEW = JJ;\]
\[KURT = JJ;\]
\[\text{IF} (\text{NR} - \text{NC}) > 1 \text{ THEN } MEAN = SS[NV + 1];\]
\[\text{IF} (\text{NR} - \text{NC}) > 2 \text{ THEN } SKEW = SS[NV + 2];\]
\[\text{IF} (\text{NR} - \text{NC}) > 3 \text{ THEN } KURT = SS[NV + 3];\]
\[SS = SS[I; NV];\]
\[VAR = \text{DIAG}(SS);\]
\[SQTIVAR = \sqrt{(\text{VECDIAG}(VAR))};\]
\[DIV = \text{DIAG}(SQTIVAR);\]
\[R = DIV * SS * DIV;\]
\[SS = I(NN, NV, \text{SEED});\]
\[SS = \text{NORMAL}(SS);\]
\[\text{IF} (\text{SKEW} = JJ \text{ AND } KURT = JJ) \text{ THEN DO;}\]
\[U = \text{ROOT}(R);\]
\[SS = SS * U;\]
\[END;\]
\[\text{ELSE DO;}\]
\[\text{PRINT "DATA ARE NOT NORMAL";}\]
\[\text{RUN FLEISH(NN, NV, SS, R, SKEW, KURT);}\]
\[\text{END;}\]
\[SS = SS * SQTIVAR + J(NN, 1, 1) * MEAN;\]
\[\text{PRINT SS;}\]
\[\text{END;}\]
\[\text{FINISH;}\]

\[\text{START MAIN;}\]
\[\text{PRINT "BEGINNING MAIN ROUTINE.";}\]
\[\text{USE A;}\]
\[\text{READ ALL INTO SS;}\]
\[\text{PRINT "DATA READ INTO SS.";}\]

\[\text{********** SPECIFY THE FOLLOWING PARAMETERS FOR READIN} \]
\[\text{SUBROUTINE} \]
\[\text{1. SS: TO KEEP THE SIMUALTED DATA MATRIX.} \]
\[\text{2. NUMBER OF CASES.} \]
\[\text{3. SEED TO GENERATE RANDOM NUMBERS.} \]

\[\text{**********} \]

\[\text{RUN READIN(SS, 10000, 123456789);}\]
\[\text{CREATE B FROM SS;}\]
\[\text{APPEND FROM SS;}\]
\[\text{FINISH;}\]

\[\text{RUN MAIN;}\]

\[\text{DATA C;}\]
\[\text{SET B;}\]
\[\text{PROC MEANS MEAN VAR SKEWNESS KURTOSIS MIN MAX;}\]
\[\text{PROC CORR;}\]
\[\text{RUN;}\]