1. Introduction

Assessing the distributional assumptions in multivariate statistics is one of the most neglected areas in modern statistics. Violation of the assumption of multivariate normality can have a serious effect on the validity of the estimates. The widely used maximum likelihood estimate, for example, is not robust with regard to non-normality (Boomsma, 1983; Harlow, 1986; Tanaka, 1984). Multivariate normality represents only an ideal, approximations of which are rarely seen in empirical data sets. Without proper tests for the distributional assumption of the observed variables, multivariate statistical estimations, like maximum likelihood, might not yield appropriate statistical statements.

There exist many different methods to test the normality of data. Shapiro and Wilk (1968) compared nine different methods in their study. Based on findings from five different sample sizes ranging from 5 to 50, they concluded that a combination of both univariate skewness and kurtosis usually provides a sensitive judgement of non-normality. The univariate skewness can be viewed as indicating how much a distribution departs from symmetry, while univariate kurtosis has been described as an indication of the peakness of a distribution. The univariate skewness and kurtosis are informative regarding the marginal distribution of a variable. However, univariate skewness and kurtosis have been described as an indication of the peakedness of a distribution. Although univariate measures of skewness and kurtosis are informative regarding the marginal distribution of a variable, it is also of interest to have information on the joint distribution of a set of variables. Mardia (1970-1974) developed multivariate measures of skewness and kurtosis that are helpful in assessing multivariate normality.

The SAS system provides the univariate skewness and kurtosis statistics in PROC UNIVARIATE. However, it does not have a procedure that can calculate the multivariate counterpart of these statistics. In this paper, we describe a PROC MATRIX program which provides Mardia-proposed multivariate skewness and kurtosis statistics. In addition, this program prints an indicator of the general tendency of a observation to be an outlier in multidimensional space. and eliminates it. The observation that makes the largest contribution to the multivariate kurtosis measure is picked and deleted. After that, this program calculates the multivariate skewness and kurtosis again, based on the remaining data. The soon-to-be extinct PROC MATRIX of SAS Version 5 is the major tool used to develop this program. The code can be easily converted to SAS/IML code in the future.

2. Multivariate Expansion of Skewness and Kurtosis

Let \( X'_1 = (X'_1, \ldots, X'_p) \) \((i=1,2,\ldots,n)\) be a random sample of size \( n \) from a \( p \)-variate population with random vector \( X'_1 = (X'_1, \ldots, X'_p) \). Let \( \tilde{X}' = (\tilde{X}_1, \ldots, \tilde{X}_p) \) and \( S' = (S'_{ij}) \) denote the sample mean vector and the covariance matrix respectively. Then the multivariate skewness, denoted by \( b_{p}' \), may be obtained by considering the canonical correlations between \( X \) and \( S \) (Hardia, 1970, P.520). The formula for multivariate skewness for a \( p \)-variate random sample of size \( n \) is:

\[
b_{p}' = \frac{1}{n(n-1)(n-2)} \sum_{j=1}^{n} \frac{1}{n-2} \left( \frac{1}{n-2} \right)^2 \frac{1}{n-2} \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)^2 \left( \frac{1}{n-2} \right)
ed in the sample by:

\[ B = \frac{(\beta r \cdot (p^2 + p + 2)(n-1)(n+1))}{(8p^2 + 2p + n)} \]

where \( B \) is distributed as a normal variate with a mean of zero and a variance of 1.0, in a large sample. Since the expected value of \( B \) is equal to zero under multivariate normality, large or small values for \( B \) are indicative of deviation from normality in the sample.

3. Detection of Outliers

In some situations the scores for an individual case may represent gross recording errors or some other anomaly that makes the scores quite unrepresentative of the population being studied. Those outlying observations are considered the cause of the non-normality. If there is an outlier, the tail of the distribution is heavier than that of the normal distribution, and thus the value for kurtosis becomes greater than zero. Hence, the further the observation departs from the centroid, the larger the impact that observation has on calculating kurtosis. The observation with the largest contribution to the measure of multivariate kurtosis can be regarded as the outlier.

The normalized estimate of \( b_p \), equation (4), is the mean across all instances of these case contributions. Thus if a case is an outlier, its contribution will be very large relative to the other cases. Eliminating such a case should have the effect of reducing the normalized multivariate kurtosis measure. However, the exact effect is, to some extent, difficult to predict because the mean and covariance matrix that are the building blocks of the statistics are themselves changed by the elimination of an outlier (Bentler, 1985).

In this program, the deviant observations that make the largest contribution to the kurtosis measure are automatically deleted one by one until the test statistics for multivariate kurtosis reach a certain cut-off point (alpha=.05). This procedure represents a kind of shotgun approach to handle the outliers. Whether one decides to delete or to keep these observations in subsequent computer runs should be based on other criteria in addition to the statistical consideration.

4. Example

The PROC MATRIX program contains the following five subroutines:

1) STATS: for calculating the basic statistics.
2) SKKT: for calculating the multivariate skewness and kurtosis.
3) MULT: for calculating the probability level of multivariate skewness and kurtosis based on Mardia-proposed criteria.
4) OUTLIER: for picking the outlier observation.
5) NEW_DAT: for eliminating the outlier and creating a new data set without the outlier.

The subroutines are called using LINK statements and the output is controlled by PRINT commands. Thus more concise output can be generated by deleting some PRINT statements.

5. Remarks

It has been suggested (e.g., Browne, 1982) that large sample sizes may be needed when fourth-order moments (i.e., kurtosis) are estimated. Unfortunately, there is no rule of thumb to choose the appropriate sample size in this situation. As a rough estimate, at least, four or five hundred observations might be required for securing the stable estimation of kurtosis. In her extensive Monte Carlo study, Harlow reported that a sample size of four hundred is quite reasonable to get stable kurtosis measures (Harlow, 1985). However, problems with using large sample size is that even small deviations from normality are liable to be statistically significant. It is possible that the distortion caused by small departures from normality may be relatively trivial as well, so that little might be lost in retaining the assumption of normality.

If multinormality is not a reasonable description of the distribution of variables (i.e., there is excess multivariate kurtosis), then normal theory procedures, such as maximum likelihood estimates, are not valid in covariance structure models due to misspecification. Therefore, corrections to the normal theory maximum likelihood ratio test statistic (Browne, 1982). Mardia's measure of multivariate kurtosis, \( b_p \), has been used in this correction (Browne, 1982, p. 85).
Therefore, the Mardia-based multivariate kurtosis might play an important role in robustifying the normal theory estimates in covariance structure models.

Acknowledgement:

The authors would like to thank Seung-joon Oh at Syracuse University for his helpful support and suggestions.

SAS and SAS/IML are registered trademarks of SAS Institute Inc., Cary, NC, USA

REFERENCES


PROC MATRIX CODE

DATA SAMPLE;
INPUT .... data input format ......;
CARDS;
..... raw data lines ........

PROC MATRIX FUZZ FW=8;
*-----------------------------------------------------------------*
MATRIX CODE TO COMPUTE MULTIVARIATE SKEWNESS AND KURTOSIS.
MODELED AFTER K.V. MARDIA'S ARTICLE: BIOMETRIKA(1970), VOL.57,
PP. 519--530.
FORTAN PROGRAM THAT CAN CALCULATE THE SAME MARDIA-BASED MULTI-
VARIATE SKEWNESS AND KURTOSIS CAN BE FOUND IN APPLIED STATISTICS,
VOL. 24, NO2 PP. 262-265: ALGORITHM AS94.
CAUTION!! THE SAMPLE SIZE SHOULD BE VERY LARGE.
*-----------------------------------------------------------------*;
FETCH X DATA=SAMPLE COLNAME=ITEMNAME;
*----------------------------------------------------------*
DATA FROM "THEORY AND IMPLEMENTATION OF EQS: A STRUCTURAL
EQUATIONS PROGRAM," BY P.M. BENTLER, P. 105
*----------------------------------------------------------*;
N=NROW(X) ;
TEMP='DEV KURT';
ITEMNAME=ITEMNAME| |ITEMNAME;
NITEM=NCOL(X);
REPEAT:
LINK STATS;
*BASIC DESCRIPTIVE STATISTICS;
NOTE STATISTICS FOR VARIABLES;
PRINT STATS ROWNAME=ITEMNAME COLNAME=STATNAME;
NOTE COVARIANCE MATRIX;
PRINT COV_CR COLNAME=ITEMNAME ROWNAME=ITEMNAME;
LINK SK KT;
*CALCULATING MULTIVARIATE SKEWNESS AND KURTOSIS;
LINK MULT;
NOTE MULTIVARIATE SKEWNESS AND KURTOSIS WITH HYPOTHESIS TESTING;
NOTE HO FOR SKEWNESS: MULT SKEW=0;
NOTE HO FOR KURTOSIS: NOR KURT=0;
NOTE (NORMALIZED KURTOSIS);
PRINT FINAL ST COLNAME=NAMESTAT;
DDEV=x| |DEV;
*CONCATENATING KURTOSIS STATS WITH DATA;
NOTE ORIGINAL DATA WITH KURTOSIS STATS;
PRINT DDEV COLNAME=ITEMNAME;
LINK OUTLIER;
*FINDING OUT THE OUTLIER BASED ON KURTOSIS;
NOTE OBSERVATION WITH LARGEST CONTRIBUTION TO MULTIVARIATE KURTOSIS;
FREE OUTLIER BIG_KURT DEV_PROP LIER DEV N KTNAME DDEV XX X_S;
LINK NEW DAT;
*NEW DATA WITHOUT OUTLIER;
NOTE NEW DATA WITHOUT OUTLIER;
PRINT NEW DAT COLNAME=ITEMNAME;
FREE X N 7
*** TESTING THE SIGNIFICANCE OF MULTIVARIATE KURTOSIS:ALPHA=.05 ***;

1181
IF PROB_KT> .05 THEN GOTO FINISH; *NOT SIGNIFICANT;
ELSE LINK NEWDATA;

FINISH: STOP;

*----------------------------------------------*
I SUBROUTINE TO CALCULATE THE BASIC STATISTICS I
*----------------------------------------------*
STATS:
NN=N*J(NITEM,1,1);NN=NN';
MEAN=X(.);;
SUM=X(+);;
COV_CR=(X'*X-SUM'*SUM/N)/(N-1);
VARIANCE=VEC_DIAG(COV_CR);VARIANCE=VARIANCE';
STD=SQRT(VARIANCE<0);;
MINIMUM=X(<);;
RANGE=MAXIMUM-MINIMUM;
STATS=NN/MEAN/STD/VARIANCE/MINIMUM/MAXIMUM/RANGE;
STATNAME='OBS' 'MEAN' 'STD DEV' 'VARIANCE' 'MINIMUM' 'MAXIMUM' 'RANGE';
RETURN;

*--------------------------------------------------*
I SUBROUTINE TO CALCULATE MULTIVARIATE SKEWNESS I
 AND KURTOSIS. I
*--------------------------------------------------*
SK_KT:
FREE COV_CR STATS NN STD VARIANCE MINIMUM MAXIMUM RANGE;
COV=(X'*X-SUM'*SUM/N)/N;
INVS=GINV(COV);
XX=X-(J(N,1,1)*MEAN);
S=X*INVS;
SK=S*XX';
SKEW=SXX';
MRD=SK/(N);;
SKK=MRD;
SKEWT=MRD/(N);;
FREE SUM MEAN SKK XX COV X_S;
C2=SKK';
B2=TRACE(C2);
FREE SK ;
KURT=B2/(N);
DEV=VECDIAG(C2);
BIG_NO=DEV<0>;
BIG_KURT=DEV<0>;;
DEV_PROP= BIG_KURT#/B2;
RETURN;

*--------------------------------------------------*
I SUBROUTINE TO CALCULATE THE PROBABILITY LEVEL OF I
SKEWNESS AND KURTOSIS BASED ON MARDIA-PROPOSED I
CRITERIA. I
*--------------------------------------------------*
MULT:
A=(N*SKW##/6;
DF_SK=(NITEM*(NITEM+1)*(NITEM+2))/6;
PROB_SK=1-PROBSCHI(A,DF_SK);
BETA=(NITEM*(NITEM+1))/(N+1);
K_DENOM=SQRT((B*NITEM*(NITEM+2))/N);
B=(KURT-BETA)/K_DENOM;
PROB_KT=1-PROBNORM(B);
NAMESTAT='SKEWNESS' Prob SK 'DF SK' 'NOR_KURT' 'Prob KT';
FINAL STAT=SKW##|Prob SK |DF SK |B|Prob KT;
RETURN;
SUBROUTINE TO PICK-UP THE OUTLIER OBSERVATION

OUTLIER:
LIER=DDEV(BIG NO.);
DEV N='TOT KURT' 'DEV_KURT' 'PROP';
OBS N='OBS NO';
KTNAME='OBS N' | ITEMNAME | DEV N;
OUTLIER=BIG_NO | LIER | B2 | DEV_PROP;
RETURN;

SUBROUTINE TO ELIMINATE THE OUTLIER OBSERVATION

NEW DAT:
INDEX A=BIG NO-1;
INDEX B=BIG NO+1;
MAT A=X(1:INDEX A);
MAT B=X(INDEX B:N);
NEW DAT=MAT_A/MAT_B;
RETURN;

INITIALIZING NEWDATA WITHOUT OUTLIER

NEWDATA:
X=NEW DAT;
N=NROW(X);
FREE NEW DAT;
GOTO REPEAT;
RETURN;