42a

*Multivariate Behavioral Research*, 1988, 23, 131–132
*Computer Software*

# UTRANS and MTRANS: Marginal and Joint Box-Cox Transformations of Multivariate Data to "Near" Normality

Hamparsum Bozdogan and Donald E. Ramirez
University of Virginia

## *UTRANS: Univariate Transformations*

UTRANS computes the mean vector, maximum likelihood covariance matrix, correlation matrix, and inverse of the covariance matrix. It computes multivariate measures of *skewness* and *kurtosis* and tests the significance of both. It gives the correlation between the original variables and their corresponding normal scores, and provides the Q-Q plots to assess the normality of the untransformed observations. It also computes the univariate skewness and kurtosis for each of the variables. UTRANS standardizes to data to have mean 10, standard deviation 1.5, and positive skewness. This is required to make the data nonnegative, to allow use of the Box-Cox (1964) transformations. UTRANS transforms the standardized data by marginal Box-Cox transformations to "nearly" normal data, provides $L_{max}$ ($\lambda_i$) and the corresponding P-value for the transformation for each of the variables. Later, UTRANS re-computes the multivariate skewness and kurtosis and tests the effect of the transformation for comparative inference as a rough guide. Again, the correlations between the transformed variables and their normal scores are computed, and Q-Q plots of these are provided to assess the "near" normality of the transformed data graphically on each of the variables.

## *MTRANS: Multivariate Transformations*

MTRANS transforms multivariate data to "nearly" normal data by a joint Box-Cox transformation using Akaike's (1973) Information Criterion (AIC). The Box-Cox values can be computed by using either the (1) *Grid Search* (GS) or (2) *Quasi-Newton* (QN) method. If the user selects the GS algorithm, then the "best" $\lambda$ (transformation parameter) is chosen for each of the variables by the minimum value of AIC marginally. Further, the difference of AIC (i.e., $\Delta$AIC) between the

Correspondence should be addressed to either author at the Department of Mathematics, University of Virginia, Charlottesville, VA, U.S.A. 22903.

original and transformed data is computed to detect variables which are "nearly" normal. After the univariate estimates for $\lambda_i$ are obtained, the GS algorithm uses those values and searches other alternative $\lambda_i$'s until no further improvement of the AIC is achieved.

If the user selects the QN algorithm, then one may enter the optimal univariate $\lambda$ vector as input, or, alternatively, provide arbitrary values. The "best" $\lambda$ vector is found by the QN method, and the minimum AIC is given along with $\Delta$AIC between the original and the transformed data, to test the significance of multivariate normality of the transformed data. The transformed data from MTRANS can be assessed for multivariate normality by measures of skewness and kurtosis, using UTRANS.

*Limitations and Availability*

Both UTRANS and MTRANS have been written in ANSI FORTRAN 77, and both require IMSL subroutines. The programs have been used successfully by the authors on a Prime 550/75/9955 computer at the University of Virginia. A copy of these programs (with a sample run for each) is available from either of the authors on an IBM compatible floppy diskette, on an as-is basis, for $50.

## References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Eds.), *Second International Symposium on Information Theory*. Budapest: Academiai Kiado.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society (B), 26,* 211–252.