DOCUMENT RESUME

ED 360 348 TM 020 223

AUTHOR	Kirisci, Levent; Hsu, Tse-Chi
TITLE	The Effect of the Multivariate Box-Cox Transformation on the Power of MANOVA.
PUB DATE	Apr 93
NOTE	19p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Computer Simulation; Equations (Mathematics); *Mathematical Models; *Matrices; *Multivariate Analysis; Sample Size; *Statistical Distributions; *Transformations (Mathematics)
IDENTIFIERS	*Box Cox Transformations; Nonnormal Distributions; *Power (Statistics)

ABSTRACT

Most of the multivariate statistical techniques rely on the assumption of multivariate normality. The effects of non-normality on multivariate tests are assumed to be negligible when variance-covariance matrices and sample sizes are equal. Therefore, in practice, investigators do not usually attempt to remove non-normality. In this simulation study, the effects of non-normality on skewed multivariate data in terms of power were examined by manipulating the factors such as distribution, sample size, number of variables, and variance-covariance matrix. The number of replications was set to 500, and sample sizes of 10, 15, and 20 were used, with 2 sets of variables, and 2 variance-covariance matrices. The multivariate Box-Cox transformation was applied to remove non-normality. The power of multivariate analysis of variance (MANOVA) was then calculated after the transformation. The results were compared with the power calculated before the multivariate Box-Cox transformation was applied. In conclusion, even when variance-covariance matrices and sample sizes were equal, small to moderate increases in power were observed. (Author/SLD)

The Effect of The Multivariate Box-Cox Transformation on The Power of MANOVA

by

Levent Kirisci, Ph.D.

University of Pittsburgh

and

Tse-Chi Hsu, Ph.D.

University of Pittsburgh

Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, April 1993.

Abstract

Most of the multivariate statistical techniques rely on the assumption of multivariate normality. The effects of nonnormality on multivariate tests are assumed to be negligible when variance-covariance matrices and sample sizes equal. Therefore, in practice, investigators usually do not attempt to remove nonnormality.

In this simulation study, the effects of nonnormality on skewed multivariate data in terms of power were examined by manipulating the factors such as distribution, sample size, number of variables, and variance-covariance matrix. The multivariate Box-Cox transformation was applied to remove nonnormality. The power of MANOVA was then calculated after the transformation. The results were compared with the power calculated before the multivariate Box-Cox transformation applied. In conclusion, even variance-covariance matrices and sample sizes were equal, small to moderate increases in power were observed.

Introduction

Most of the univariate and multivariate statistical techniques assume additivity, homogeneity of variances, normality, and independence of observations. In practice, it is difficult to attain all these assumptions simultaneously. Therefore, most investigators face the issue of assumption violations in their research. To counter the issue, Tukey (1977) proposed two basic solutions to deal with possible violations of the first three assumptions: (a) to employ a non-linear or linear transformation to data to meet the assumptions, or (b) to develop a new statistical technique that fits data better.

The first option generally gets higher acceptance in practice because the second option involves greater investment of time and effort. Therefore, numerous transformation techniques have been extensively studied and reviewed by several researchers for univariate cases (eg., Hoyle, 1973). Furthermore, the effects of violations on normality, additivity, and homogeneity of variances were also investigated (e.g., Box, 1954; Tiku, 1971; Harwell, Rubinstein, Hayes & Olds, 1992). Fortunately, the tests developed for univariate linear models are quite robust to violations of assumptions in most of the cases (Glass, Peckman & Sanders, 1972). However, the violations of assumptions in multivariate case have not been studied as extensively as those in univariate case. Mardia (1971) examined the effects of nonnormality on multivariate regression tests and one-way MANOVA. He concluded that when variance-covariance matrices and group sample sizes were equal, the effects of nonnormality on multivariate tests were considerably negligible. However, Mardia (1971) did not investigate the robustness of the multivariate general linear models in depth for different distributional assumptions. Therefore, the effects of nonnormality on the multivariate general linear models deserve a further study.

The purpose of this simulation study was to investigate the effectiveness of the multivariate Box-Cox transformation in normalizing the distribution of multivariate data and its effect on the power of MANOVA under various sample sizes, number of variables, variance-covariance structures, and distributional assumptions.

Theoretical Perspective

Box and Cox (1964) have suggested a family of transformation to normalize observations, to stabilize variance, and linearize the relationship between dependent and independent variables in regression. The notable examples of this family of transformations are (a) square-root transformation to stabilize variance and to remove non-normality, (b) cube-root transformation to remove nonnormality, and (c) logarithmic transformation to stabilize variance and to remove nonnormality.

Box and Cox considered a family of transformations for $x>0$[1]

$$y(\lambda) = \begin{cases} \dfrac{x^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0 \\ \log_e x & \text{if } \lambda = 0 \end{cases}$$

to simultaneously satisfy all three assumptions. The coefficient λ can be estimated by using the maximum likelihood method. The maximum likelihood estimate of λ maximizes the likelihood function $L(\lambda)$. Furthermore, to test whether maximum likelihood estimate λ is statistically equal to 1, that denotes a normality, the following likelihood ratio test have been proposed:

[1]The Box-Cox power transformation technique can also be used for both positive and negative numbers if $y(\lambda)=(x-\zeta)^{\lambda}$ is replaced for x^{λ}. The joint maximum likelihood estimates of ζ and λ then will be estimated by maximizing Likelihood function of (ζ,λ).

$$2\{L_{max}(\lambda)-L_{max}(1)\} \le \chi^2{}_1(\alpha)$$

where $\chi^2{}_1(\alpha)$ denote the upper 100α% point of χ^2 with 1 degree of freedom. To avoid the correlation effect between λ and θ where $E(y(\lambda))=X\theta$, X is a vector of observations of x's and θ is an unknown parameter, Box and Cox (1982) later modified the above transformation and proposed to use the following formula:

$$y(\lambda) = \frac{x^{\lambda}-1}{\dot{x}^{(\lambda-1)}} \quad \text{if } \lambda\neq 0$$

$$= \dot{x}\log_e x \quad \text{if } \lambda=0$$

where $\dot{x}$ is a geometric mean of all observations. Box, Hunter and Hunter (1978) provided an example how to use the Box-Cox transformation for univariate case with a graphical demonstration. Since it requires an extensive calculations, Hinkley (1977) and Emerson and Stoto (1982) suggested to estimate λ by using transformation plot for symmetrizing and straightening the relationship between dependent and independent variables. Hines and Hines (1987) also presented a chart to calculate λ approximately.

Andrews, Gnanadesikan and Warner(1973) and Hernandez and Johnson(1980) generalized the Box-Cox transformation to multivariate data and its significance test for λ=1 or $(\lambda_1,..,\lambda_p)=(1,..,1)$ where p is a number of variables. For each given λ, the transformation can be defined as

$$y_{ij}(\lambda) = \frac{x_{ij}{}^{\lambda j}-1}{\lambda_j} \quad \text{if } \lambda_j\neq 0$$

$$= \log_e x_{ij} \quad \text{if } \lambda_j=0$$

where i=1,..,n (n=number of observations), j=1,..,p (p=number of variables) and

$x_{ij}>0$. The maximum likelihood estimate of λ is calculated by maximizing

$$L(\lambda) = (-n/2)\log|\Sigma|+\{\Sigma_{j=1,p}(\lambda_j-1)\Sigma_{i=1,n}\log y_{ij}\}$$

where Σ can be replaced by its maximum likelihood estimate. The corresponding significance test then is

$$2\{L_{max}(\lambda)-L_{max}(1)\} \leq \chi^2{}_p(\alpha) \qquad (1)$$

that is the same as univariate case except χ^2 distribution has p degrees of freedom, where p corresponds to the number of variables. Rode and Chinchilli (1988) suggested to use the Newton-Raphson iterative algorithm to obtain the maximum likelihood estimate of λ.

Method

Data were generated from two independent populations with equal variance-covariance matrices and sample sizes. However, their corresponding group mean vectors were assumed to be different by a $.5\sigma$ (medium effect size) (Cohen, 1988) such that $\mu_{1(i)}-\mu_{2(i)}=.5\sigma_{ii}$ where 1 and 2 denoted group membership and i stood for a variable i. Type I error rate was set to 0.05 throughout the study. The number of replications were limited to 500.

The following factors were manipulated:

(a) Three sample sizes were examined, $n_1=n_2=10$, 15, 20.

(b) Two sets of variables were employed, p=2 or p=3.

(c) Two variance-covariance matrices were specified:

Case 1. $\Sigma_1=\Sigma_2=\Sigma$ where variables were uncorrelated, $\rho_{ij}=0$, $i\neq j$, for all i and j.

Case 2. $\Sigma_1=\Sigma_2=\Sigma$ where variables were correlated such that $\rho_{12}=.60$, $\rho_{13}=.55$, and $\rho_{23}=.35$.

(d) The following combinations of distributions were specified:

Case 1. When p=2, variable 1 was highly skewed and moderately leptokurtic (skewness=1.75 and kurtosis=4.00) and variable 2 was moderately skewed and slightly leptokurtic (skewness=1.00 and kurtosis=1.00).

Case 2. When p=3, variable 1 and variable 2 were specified as in case 1. Variable 3 was highly skewed and highly leptokurtic (skewness=2.25 and kurtosis=6.50).

Multivariate nonnormal and skewed data were generated by using GENRAW2 (Joreskog and Sorbom, 1989) and ITMGNR (Kirisci, 1992) computer programs. The maximum likelihood estimates of λ were obtained by using a Fortran IV computer program written by the authors. The maximum likelihood estimate of λ was calculated 500 times for each combination. The average value along with its standard deviation score were presented. Furthermore, the significance level of the likelihood ratio test statistic for normality (λ=1, see equation 1) was computed and its average significance level was reported. Power and noncentrality parameter for MANOVA were computed by using SPSSX and their average scores were presented. Power and noncentrality were calculated by utilizing the following formulas (Morrison, 1976):

$$\text{Power} = 1-\beta(\delta^2) = \Pr(F' > F_{\alpha;p,(n1+n2-1)-p+1}),$$

where F' is a noncentral F distribution and

$$\text{Noncentrality parameter} = \delta^2 = (n_1+n_2)(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2).$$

where Σ^{-1} is a inverse matrix of variance-covariance matrix Σ.

Finally, chi-square probability plots of squared radii, that was proposed by Andrew, Gnanadesikan and Warner (1971, 1973), were drawn by AXUM (TriMetrix, 1992). To draw a chi-square probability plot, Mahalonobis distances (D_j) were calculated and ranked from the smallest to the largest. Theoretical chi-square scores were obtained for each 100(j-.5)/n percentile from CHIDF (IMSL, 1989), where n was a sample size and j varied between 1 and n. As a final step, points (D_j, 100(j-.5)/n)) were plotted in two-dimensional space. The plot suggests normality, if all points lie closer to a 45-degree straight line.

Results

In this simulation study, the effects of nonnormality on skewed multivariate data in terms of power were examined. The factors such as distribution, sample size, and number of variables were manipulated in order to decide whether normality assumption was crucial in MANOVA when the number of observations per variable ranged between moderate values. Since the variance-covariance matrices and sample sizes were assumed equal throughout the study, we anticipated that improvements in power after the transformation should be relatively small, if MANOVA is robust to the violation of nonnormality assumption.

According to the results summarized in Table 1, the multivariate Box-Cox transformation had a notable effect on the power of the test as well as on the noncentrality parameter. The smallest increase in power (.09%) was observed when 2 variables were correlated and the sample size was 15 for each group. The highest increase (26.1%) was attained when 2 variables were uncorrelated and the sample size was 20 for each group.

Please insert Table 1 here

To place a broader perspective the effects of the multivariate Box-Cox transformation on the power of MANOVA, it may be useful to examine the maximum likelihood estimates of λ. Correlated and uncorrelated cases should be examined separately, since the maximum likelihood estimates can be affected if variables are correlated. When the number of variables was 2 and they were uncorrelated, the first variable that was specified as highly skewed and moderately leptokurtic took higher values than those of the second variable that was moderately skewed and slightly leptokurtic. The maximum likelihood estimates corresponding to the first variable suggested that, on the average, the maximum likelihood estimate of λ was .16. For the second variable, the values was closer to zero. In practice, this may suggest that a logarithmic transformation was appropriate for the second variable that was moderately skewed and slightly leptokurtic. The results of the likelihood ratio test for normality and their average significance level indicated that when sample sizes was 10, the transformation for normality was successful (nonsignificant result) at .05 level. When sample size increased from 10 to 20, the average significance level dropped to 0.0001. Since χ^2-test are very sensitive to number of observations and outliers, this is ususally the case in goodness-of-fit tests. When the number of variables were 3 and uncorrelated, the maximum likelihood estimates of λ were, on the average, 0.10 for the first variable, 0.20 for the second variable, and 0.56 for the third variable that was highly skewed and highly leptokurtic. This suggested that to remove nonnormality, λ in Box-Cox transformation should be set to 0.10 for the first variable, 0.20 for the second variable and 0.55 for the third variable (square-root transformation). The average significance levels followed the same pattern as before; it dropped from 0.0034 to 0.0000.

When the number of variables were 2 and correlated, the maximum likelihood estimates of λ corresponding to the first variable were closer to 0 that denoted a logarithmic transformation . For the second variable, it was closer to 0.30, a cube-root transformation was appropriate. The significance levels changed from 0.0281 to 0.0004. When the number of variables were 3, the maximum likelihood estimates of λ corresponding to the first variable suggested a square-root transformation. The second variable was transformed to remove nonnormality by setting λ=1.17. The third variable required a logarithmic transformation.

The effects of Box-Cox tran formation on multivariate data can be observed by examining some of the selected chi-square probability plots that were presented in Figure 1, Figure 2, and Figure 3. As can be noted, if multivariate normality holds, all the points to be expected lie on the 45-degree straight line. By considering that, it is possible to visualize the effects of the multivariate Box-Cox transformation on normality.

Please insert Figure 1, Figure 2 and Figure 3 here

Conclusion

In educational research and as well as in behavioral sciences, investigators heavily rely on multivariate statistical techniques in analyzing multivariate data with complex structures. Multivariate normality is one of the key assumptions that underlies much of the classical multivariate statistical techniques. Therefore, to meet this assumption gains greater importance in data analysis.

In this study, our attention focused on the violation of normality assumption for multivariate data and its effect on power. The multivariate Box-Cox transformation technique was applied to nonnormal and skewed multivariate data. Power and noncentrality parameter were calculated for MANOVA in order to show the effects of the multivariate Box-Cox transformation. Lastly, a chi-square probability plot of squared radii was employed to transform multivariate data into unidimensional space to view how successful the multivariate Box-Cox transformation technique in achieving normality.

Although the multivariate Box-Cox transformation is laborious and time consuming, its drawbacks are only a small price to pay for the benefits that can be obtained from the information we gain. A computerized literature survey of articles in psychology and behavioral sciences since 1988 using PSYCHINFO, produced by the American Psychological Association, shows that transformation of multivariate data for normality has almost never been employed in practice. Instead, most researchers assume that under the assumptions of equal variance-covariance matrices and of equal group sample sizes, the violation of normality has a small effect on the power of MANOVA. However, the results of this study showed that even if two groups had same variance-covariance matrices and equal sample sizes, small to moderate increases in power were observed. Therefore, It would be highly advisable to apply the multivariate Box-Cox transformation technique to multivariate data to remove non-normality.

References

Andrews, D.F., Gnanadesikan, R., & Warner, J.L. (1971). Transformations of Multivariate Data. Biometrics, 27, 825-840.

Andrews, D.F., Gnanadesikan, R., & Warner, J.L. (1973). Methods for Assessing Multivariate Normality, In P.R. Krishnaiah(Ed.) Multivariate Analysis-III (pp. 95-116). New York: Academic Press.

Box, G.E.P. (1954). Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems I: Effects of Inequality of Variance in the One-Way Classification. Annals of Mathematical Statistics, 25,290-302.

Box, G.E.P. & Cox, D.R. (1964). An Analysis of Transformations. Journal of Royal Statistical Society B, 26, 211-252.

Box, G.E.P. & Cox, D.R. (1982). An Analysis of Transformations Revisited, Rebutted. Journal of the American Statistical Association, 77, 209-210.

Box, G.E.P., Hunter, W.G., & Hunter, J.S. (1978). Statistics for Experimenters. New york: Wiley.

Cohen, J. (1988). Statistical Power Analysis For the Behavioral Sciences. Hillsdale, N.J.: Erlbaum Associates.

Emerson, J.D. & Stoto, M.A. (1982). Exploratory Methods for Choosing Power Transformations. Journal of the American Statistical Association, 77, 103-108.

Glass, G.V., Peckman, P.D., & Sanders, J.R. (1972). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analysis of Variance and Covariance. Review of Educational Research, 42, 237-288.

Harwell, M., Rubinstein, E., Hayes, W., & Olds, C. (1992). Summarizing Monte-Carlo Results in Methodological Research: The One- and Two-Factor Fixed Effects ANCOVA Cases. Journal of Educational Statistics, 17, 4, 315-339.

Hernandez, F. & Johnson, R.A. (1980). The Large-Sample Behavior of Transformations to Normality. Journal of the American Statistical Association, 75, 855-861.

Hines, W.G.S. & Hines, R.J.O. (1987). Quick Graphical Power-Law Transformation Selection. The American Statistician, 41, 21-24.

Hinkley, D. (1977). On Quick Choice of Power Transformation. Applied Statistician, 26, 67-69.

Hoyle, M.H. (1973). Transformations: An Introduction and a Bibliography. International Statistical Review, 41(2), 203-223.

IMSL. (1989). User's Manual: Stat/Library. Houston: Author.

Joreskog, K.G. & Sorbom, D. (1989). Lisrel 7: User's Reference Guide. Scientific Software, Inc.

Kirisci, L. (1992). ITMGNR: A Fortran IV Computer Program to Generate Item Response Vectors. Unpublished manuscript. University of Pittsburgh, Pittsburgh.

Mardia, K.V. (1971). The effect of non-normality on some multivariate tests and robustness to non-normality in the linear model. Biometrika, 58(1), 105-121.

Morrison, D.F. (1976). Multivariate Statistical Methods. Tokyo: McGraw-Hill Kogakusha, Ltd.

Rode, R. & Chinchilli, V.M. (1988). The Use of Box-Cox Transformations in the Development of Multivariate Tolerance Regions With Applications to Clinical Chemistry. The American Statistician, 42, 23-30.

Tiku, M.L. (1971). Power Function of the F-Test Under Non-Normal Situations. Journal of the American Statistical Association, 66, 913-916.

TriMetrix. (1992). AXUM: Technical Graphics and Data Analysis. Seattle: Author.

Tukey, J.W. (1977). Exploratory Data Analysis. Reading: Addison-Wesley.

Table 1.

Descriptive statistics of Power and Noncentrality parameter for MANOVA Before and After the Multivariate Box-Cox Transformation and the Maximum Likelihood Estimates of λ

	Maximum Likelihood Estimates $\hat{\lambda}_1$ Mean(sd)	Maximum Likelihood Estimates $\hat{\lambda}_2$ Mean(sd)	Likelihood Ratio Test for Normality (λ=1)[4] Significance	Before Box-Cox Transformation Power[5](δ^2[3])	After Box-Cox Transformation Power(δ^2)	% of Increase in Power
p=2						
Uncorrelated						
($n_1=n_2=10$)	.16(.07)	.06(.07)	.0241	.35(3.95)	.38(4.29)	8.6
($n_1=n_2=15$)	.15(.07)	-.10(.05)	.0032	.42(4.43)	.49(5.45)	16.7
($n_1=n_2=20$)	.17(.06)	-.02(.05)	.0001	.46(4.86)	.58(6.42)	26.1
Correlated						
($n_1=n_2=10$)	-.02(.08)	.34(.10)	.0281	.45(5.22)	.50(6.08)	1.1
($n_1=n_2=15$)	-.05(.08)	.30(.07)	.0100	.53(5.91)	.58(6.65)	0.9
($n_1=n_2=20$)	-.10(.07)	.24(.06)	.0004	.56(6.19)	.66(7.71)	17.9

[4] Significance levels were obtained from the l'kelihood ratio test statistic for multivariate normality, $2(L_{max}(\lambda)-L_{max}(1))\leq\chi^2_p(\alpha)$.

[5] $Power=1-\beta(\delta^2)=Pr(F'>F_{\alpha;p,(n1+n2-1)-p+1})$, where F' is a non-central F distribution.

[3] Noncentrality parameter$=\delta^2=(n_1+n_2)(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)$

Table 1 (continued)

	Maximum Likelihood Estimates $\hat{\lambda}_1$ Mean(sd)	$\hat{\lambda}_2$ Mean(sd)	$\hat{\lambda}_3$ Mean(sd)	Likelihood Ratio Test for Normality (λ=1) Significance	Before Box-Cox Transformation Power(δ^2)	After Box-Cox Transformation Power(δ^2)	% of Increase in Power
p=3							
Uncorrelated							
(n_1=n_2=10)	.10(.08)	.28(.07)	.64(.09)	.0034	.36(5.09)	.39(5.66)	8.3
(n_1=n_2=15)	.10(.07)	.19(.07)	.48(.06)	.0001	.46(6.00)	.50(6.69)	8.7
(n_1=n_2=20)	.09(.05)	.12(.05)	.55(.05)	.0000	.50(6.44)	.59(7.58)	18.0
Correlated							
(n_1=n_2=10)	.58(.11)	1.15(.13)	.07(.15)	.0071	.52(7.72)	.59(9.00)	13.5
(n_1=n_2=15)	.62(.11)	1.18(.10)	-.06(.13)	.0011	.68(9.90)	.70(10.5)	2.9
(n_1=n_2=20)	.50(.10)	1.17(.08)	-.05(.09)	.0001	.69(9.63)	.79(11.95)	14.5

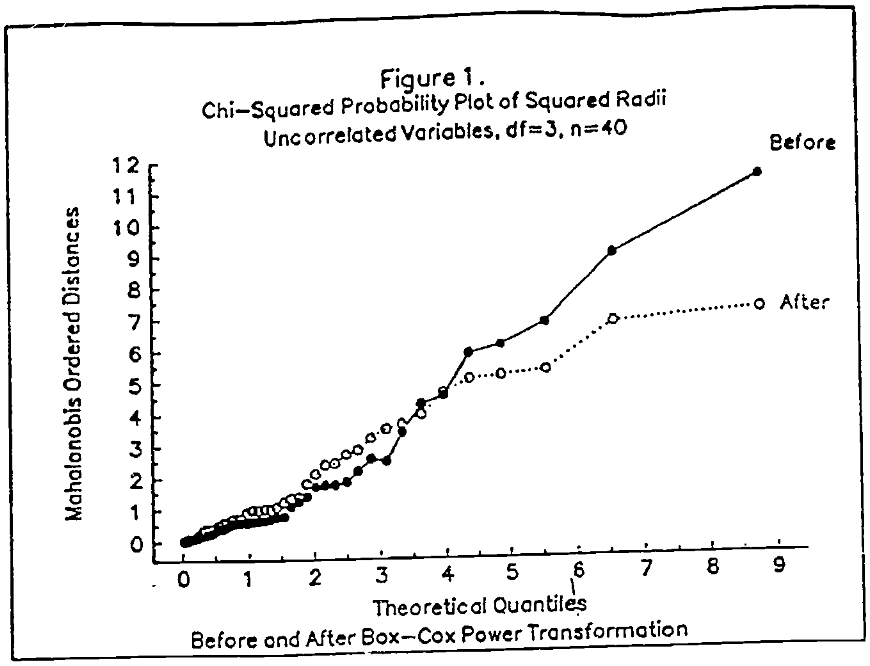
Figure 1.
Chi-Squared Probability Plot of Squared Radii
Uncorrelated Variables, df=3, n=40
Before
After
Mahalanobis Ordered Distances
Theoretical Quantiles
Before and After Box-Cox Power Transformation

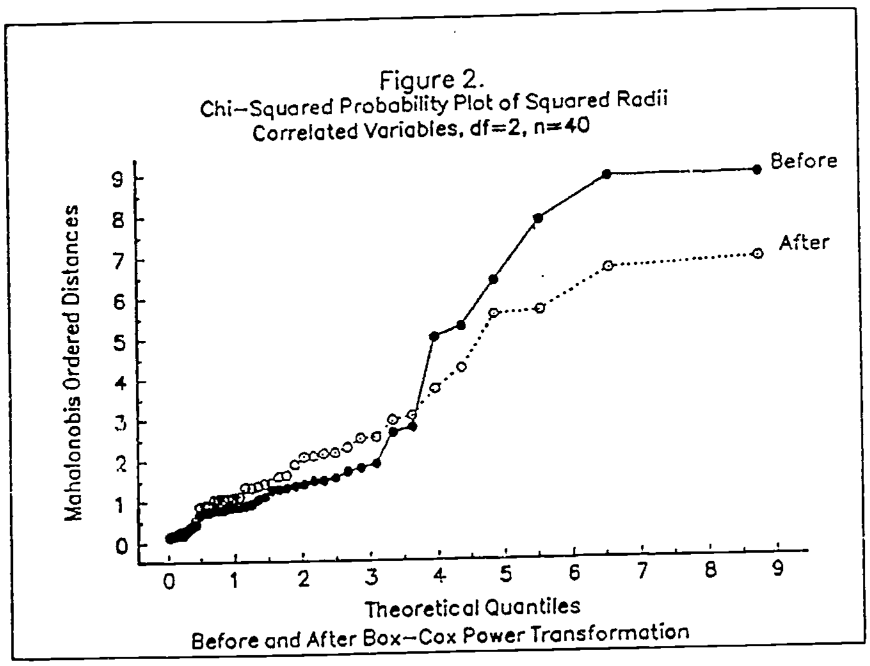
Figure 2.
Chi-Squared Probability Plot of Squared Radii
Correlated Variables, df=2, n=40
Before
After
Mahalanobis Ordered Distances
Theoretical Quantiles
Before and After Box-Cox Power Transformation

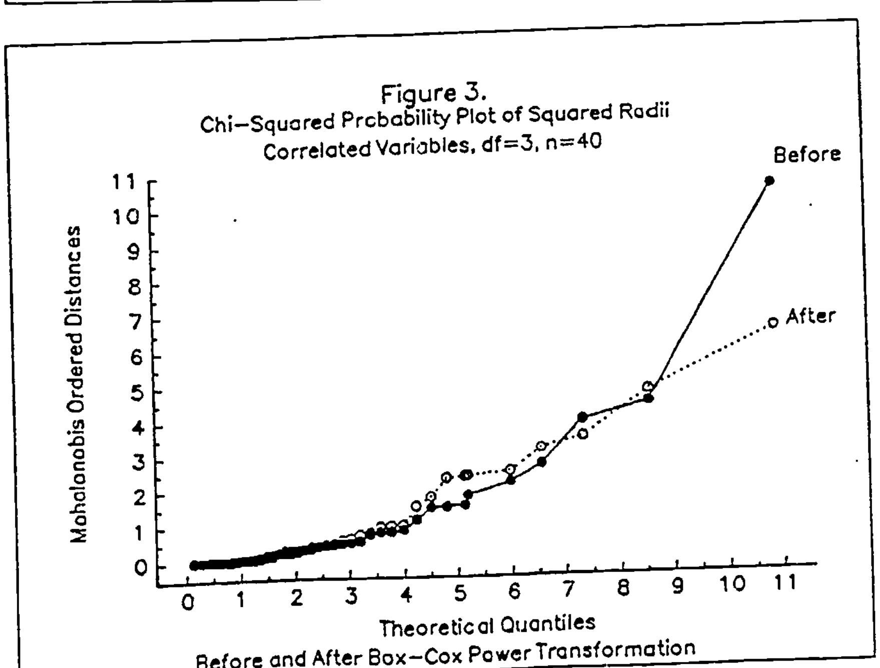
Figure 3.
Chi-Squared Probability Plot of Squared Radii
Correlated Variables, df=3, n=40
Before
After
Mahalanobis Ordered Distances
Theoretical Quantiles
Before and After Box-Cox Power Transformation