Jour. Ind. Soc. Ag. Statistics
Vol. XL, No. 2 (1988), pp. 164-68

AN APPROXIMATE METHOD FOR ASSESSING MULTIVARIATE NORMALITY

U. C. JAISWAL
Haryana Agricultural University, Hissar-125004
and
J. P. JAIN
IASRI, New Delhi-110012

(Received : December, 1987)

SUMMARY

A simple approximate method for testing multivariate normality has been proposed and compared empirically with the method due to Small [5]. This method besides being computationally more convenient is seen to compare quite favourably with the rigorous method.

Keywords : Testing for normality; Multivariate normality; Test statistic for assessing normality.

Introduction

Many of the standard statistical procedures used in biometric analysis are based on the fundamental assumption that the data conform to multivariate normal distribution. Various techniques for assessing the multivariate normality have been proposed from time to time (Healy, [2]; Mardia, [3]; Cox and Small, [1]; Small, [5]; Royston, [4]). All these tests are quite involved. In the present paper, therefore, a simple approximate method for assessing multivariate normality, which takes into account the necessary condition of univariate normality of marginal distributions has been proposed.

2. The Proposed Method

The proposed method for testing multivariate normality is an extension of the method of Snedecor and Cochran [6] for testing univariate normality by means of skewness and kurtosis coefficients.

Let corresponding to a p-dimensional random vector X, g_1 ($= \sqrt{b_1}$) be a $p \times 1$ vector of marginal coefficients of skewness with covariance matrix V_1 and g_2 ($= b_2 - 3$) be of kurtosis with covariance matrix V_2. Since the distributions of g_1 and g_2 approach normality very slowly, the distributions of quadratic forms $g_2' V^{-1} g_1$ ($i = 1, 2$) cannot be approximated well by χ^2 distribution. The coefficients of skewness and kurtosis for large samples of size n from univariate normal distribution can be transformed to standard normal variates by using their asymptotic variances, $6/n$ and $24/n$ respectively. Application of this, componentwise to marginal coefficients yields the vectors

$$Y_1 = g_1/\sqrt{V(g_1)}, \quad \text{and} \quad Y_2 = g_2/\sqrt{V(g_2)}$$

which are approximately independently distributed as $N(0, 1)$. The main diagonal elements of the covariance matrices U_1 and U_2 corresponding to Y_1 and Y_2 are, therefore, unity and the off-diagonal elements of U_1 are r_{ij}^3 and those of U_2 are r_{ij}^4, where r_{ij} is the sample correlation between the ith and jth variates in the original data (Small, [5]). The two statistics

$$Q_1 = Y_1' U_1^{-1} Y_1 \quad \text{and} \quad Q_2 = Y_2' U_2^{-1} Y_2$$

are therefore nearly independent and each is distributed approximately as χ_p^2. Hence the test statistic $Q_1 + Q_2$ can be treated as a χ^2 variable with $2p$ degrees of freedom.

That the proposed method is good enough for testing multivariate normality has been judged empirically from the closeness of results by this method with the rigorous method due to Small [5].

3. Empirical Validation of the Method

For judging empirically the effectiveness of the proposed method vis-a-vis the rigorous method of Small [5]. the data on four characters, namely, age at first calving (days), first lactation yield (kg), fat percentage and first calving interval (days) of 287 cows belonging to seven genetic groups, three of halfbreds of Hariana with Holstein-Friesian, Brown Swiss and Jersey and four of three-fourths with two exotic breeds, gene-

TABLE 1—TEST STATISTICS FOR SKEWNESS (Y_1) AND KURTOSIS (Y_2) FOR EACH OF THE FOUR TRAITS FOR SEVEN GENETIC GROUPS

Genetic Group	*No. of observations*	*Age at first calving*		*First lactation yield*		*Fat percentage*		*First calving interval*	
		Y_1	Y_2	Y_1	Y_2	Y_1	Y_2	Y_1	Y_2
$\frac{1}{2}F + \frac{1}{2}H$	50	1.369	0.506	1.631	−0.245	1.585	0.338	1.539	−0.931
$\frac{1}{2}B + \frac{1}{2}H$	30	0.098	−0.585	1.907	−0.221	−0.393	−0.891	1.572	−0.861
$\frac{1}{2}J + \frac{1}{2}H$	42	0.529	0.022	0.923	−0.718	−1.487	1.173	1.630	−0.713
$\frac{1}{2}F + \frac{1}{4}B + \frac{1}{4}H$	42	1.566	−0.319	1.238	−0.029	0.995	0.503	0.897	−1.303
$\frac{1}{2}F + \frac{1}{4}J + \frac{1}{4}H$	39	0.094	−0.795	1.874	0.367	1.101	1.570	1.409	−0.984
$\frac{1}{2}B + \frac{1}{4}F + \frac{1}{4}H$	45	1.473	−0.369	0.912	−1.224	1.211	−0.011	1.684	−0.066
$\frac{1}{2}J + \frac{1}{4}F + \frac{1}{4}H$	39	1.262	−0.460	1.876	1.001	0.828	−0.943	1.808	−0.393

rated during 1976-82 under the All India Coordinated Research Project on Cattle at Haryana Agricultural University, Hissar, were utilized.

Table 1, gives the values of the test statistics for skewness and kurtosis computed for each character separately for different genetic groups. All these values are seen to be non-significant implying thereby that the marginal distributions of all the four characters in each genetic group are nearly normal. The values of the test statistic ($Q_1 + Q_2$) computed for evaluating multivariate normality of each genetic group, by using the proposed method and the method due to Small [5] are presented in Table 2. The values of the test statistic obtained by either method are seen to be non-significant, for all the genetic groups confirming multivariate normality. Further it is seen that the values of the test statistic obtained by the proposed method are slightly lower than the corresponding values by the method due to Small indicating only minor distortion in the level of significance. The proposed method can, therefore, be used with almost equal effectiveness for testing multivariate normality by using the χ^2 table corresponding to a slightly lower level of significance. In addition the

TABLE 2—TEST STATISTICS FOR ASSESSING MULTIVARIATE NORMALITY FOR DIFFERENT GENETIC GROUPS

Genetic group	Q_1		Q_2		$Q_1 + Q_2$	
	Small's method	*Proposed method*	*Small's method*	*Proposed method*	*Small's method*	*Proposed method*
$\frac{1}{2}F + \frac{1}{2}H$	10.24	8.99	2.51	1.29	12.75	10.28
$\frac{1}{2}B + \frac{1}{2}H$	7.02	5.91	1.92	1.79	8.94	7.70
$\frac{1}{2}J + \frac{1}{2}H$	6.86	5.87	3.15	2.52	10.01	8.39
$\frac{1}{2}F + \frac{1}{4}B + \frac{1}{4}H$	6.80	5.70	4.67	2.75	11.47	8.45
$\frac{1}{2}F + \frac{1}{4}J + \frac{1}{4}H$	7.38	6.28	5.64	4.26	13.02	10.54
$\frac{1}{2}B + \frac{1}{4}F + \frac{1}{4}H$	8.18	7.06	3.19	1.84	11.37	8.90
$\frac{1}{2}J + \frac{1}{4}F + \frac{1}{4}H$	10.36	8.97	2.93	2.37	13.29	11.34

proposed method is computationally more convenient to work with whereas Small's method involves lot of iterations and assumes knowledge of values of certain parameters (γ, δ) for different pairs of values of $\sqrt{b_1}$ and b_2 which unfortunately are not available for all pairs.

REFERENCES

[1] Cox, D. R. and Small, N. J. H. (1978) : Testing multivariate normality, *Biometrika* **65** : 263-272.

[2] Healy, M. J. R. (1968) : Multivariate normal plotting, *Appl. Stat.* **17** : 157-161.

[3] Mardia, K. V. (1970) : Measures of multivariate skewness and kurtosis with applications, *Biometrika* 57 : 519-530.

[4] Royston, J. P. (1982) : An extension of Shapiro and Wilk's *W*-test for normality of large samples, *Appl. Stat.* **31** : 115-124.

[5] Small, J. J. H. (1980) : Marginal skewness and kurtosis in testing multivariate normality, *Appl. Stat.* 29 : 85-87.

[6] Snedecor, G. W. and Cochran, W. G. (1967) : *Statistical Methods*, 6th edn. Oxford and IBH Publishing Co., New Delhi.