Appl. Statist. (1983),
32, *No.* 2, *pp.* 121–133

Some Techniques for Assessing Multivarate Normality Based on the Shapiro–Wilk *W*

By J. P. ROYSTON

Medical Research Council, Harrow, UK

[Received May 1982. Revised January 1983]

SUMMARY

Shapiro and Wilk's (1965) W test is a powerful procedure for detecting departures from univariate normality. The present paper extends the application of W to testing multivariate normality, and also to Healy's (1968) test based on squared radii. Three examples illustrate the approach, and also the utility of careful scrutiny of lower-dimensional subsets of the data where otherwise unsuspected departures from normality may appear.

Keywords: MULTIVARIATE NORMALITY; W STATISTIC; SQUARED RADII; SUBSETS OF VARIATES; NORMAL PROBABILITY PLOTTING

1. INTRODUCTION

A fair number of test procedures for multivariate normality have been proposed in the literature. Cox and Small (1978) reviewed most of these recently, and I shall not cover the same ground again. Procedures may be said to concentrate either on combinations of univariate tests of normality (Small, 1979; Malkovich and Afifi, 1973), or on the geometrical properties in R^m of two or more variates taken together (Healy, 1968; Cox and Small, 1978), in particular, probability plots of squared radii from the centroid of the data in the metric defined by the sample covariance matrix (Healy, 1968; Small, 1978). Cox and Small (1978) considered departures represented by curvature in the variate-variate plots.

Certain general difficulties appear. Quite frequently, suggested test statistics have intractable null hypothesis distributions. Convincing power studies are rare, due to the large variety of possible types of departure from normality and the associated expense of Monte Carlo computer studies. Outliers, too, have unpredictable effects—for example, Campbell (1980) pointed out the sensitivity of estimated covariance matrices to outliers, and recommended robust estimation of these matrices. Plotting is a valuable aid, but geometry in m dimensions ($m > 3$) is hard to depict graphically; also, if m is large, the number of possible plots of subsets of variates multiplies rapidly, making for difficulties in interpretation (and heavy computation).

The present paper does not pretend to solve all these problems. Fortunately, however, many of the above conditions present related symptoms; for example curvature in the x_1 $v.$ x_2 plot is quite likely to manifest itself as a skew departure from 1-normality in x_1 or x_2 (or both), and a data transformation may correct matters. I shall show that Shapiro and Wilk's (1965) W test furnishes a set of univariate test statistics which show low correlation even when the parent variates are quite highly correlated, which makes the computation of a combined test fairly simple. I also apply the W test to plots of squared radii, and extend the treatment to subsets of variates. Three detailed illustrative examples are given.

Present address: Division of Computing and Statistics, MRC Clinical Research Centre, Watford Road, Harrow HA1 3UJ, UK

0035–9254/83/32121 $2.00

2. METHOD

Suppose the data to be examined for multivariate normality (or m-normality for short, where m is the number of variates) consist of a random sample, X, of n cases measured on m variates, with sample covariance matrix S. Let $v_1 < v_2 \ldots < v_n$ represent the data for one variate, arranged in increasing order of magnitude. Then Shapiro and Wilk's W is defined as

$$W = (\Sigma a_i v_i)^2 / \Sigma(v_i - \bar{v})^2,$$

where $\bar{v} = \Sigma v_i / n$ and $\mathbf{a}$ are the best linear unbiased coefficients of Sarhan and Greenberg (1956). Royston (1982) showed that W could be transformed to an approximately standard normal variate, z, under the hypothesis that the (unordered) v's came from a normal distribution with unspecified mean and variance: $z = ((1 - W)^\lambda - \mu)/\sigma$. λ, μ and σ were all functions of n, for which polynomial formulae were provided. Large positive values of z are evidence of non-normality.

Suppose $z_1, \ldots, z_m$ have been obtained as above, with z_i based on the ith component of each of the n observations. Define

$$k_i = \{\Phi^{-1} [\tfrac{1}{2} \Phi(-z_i)]\}^2, \quad i = 1, \ldots, m, \tag{1}$$

where

$$\Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} \exp(-\tfrac{1}{2}t^2)\, dt,$$

and note that

(a) k_i will be large when z_i is large positive (variate i showing signs of non-normality);
(b) k_i will tend to zero for large negative z_i (no departure from normality);
(c) $k_i \sim \chi_1^2$ individually.

Let $c_{ij} = \text{corr}(k_i, k_j)$ be the correlation matrix of k, and define $G = \Sigma_1^m k_i / m$. Clearly, if the source data X were uncorrelated, the corresponding W, z and k values would also be uncorrelated.

Then $G \sim \chi_m^2 / m$, and G provides a test of multivariate normality based on combined univariate (transformed) W statistics. If the variates were perfectly correlated, $k_1 = k_2 = \ldots = k_m$ and $G \sim \chi_1^2$. For intermediate parent correlations, a natural approximation to the distribution of G is $G \sim \chi_e^2 / e$, where e is the "equivalent degrees of freedom" (not necessarily an integer) derived as follows from the first two moments of G:

$$E(G) = 1,$$

$$m^2 \text{ var}(G) = 2m + \sum_{i \neq j}^{m} \sum \text{cov}(k_i, k_j).$$

Since

$$\text{var}(k_i) = 2, \; c_{ij} = \text{corr}(k_i, k_j) = \text{cov}(k_i, k_j)/\sqrt{(2 \times 2)}:$$

$$m^2 \text{ var}(G) = 2m + 2 \sum_{i \neq j} \sum c_{ij},$$

$$\text{var}(G) = \frac{2}{m} + \frac{2}{m} \sum \sum c_{ij}$$

Now $\text{var}(\chi_e^2 / e) = 2e/e^2 = 2/e$, so equating second moments:

$$e = \frac{2}{\operatorname{var}(G)} = \frac{1}{\frac{1}{m} + \frac{1}{m^2}\Sigma\Sigma c_{ij}}$$

$$= \frac{m}{1 + (m-1)\bar{c}},$$

where $\bar{c} = \Sigma\Sigma c_{ij}/(m^2 - m)$, noting that $\{c_{ij}\}_{i \neq j}$ has $(m^2 - m)$ elements. As a final step, set

$$H = eG = (e/m) \sum_{1}^{m} k_i \mathrel{\dot{\sim}} \chi_e^2 \tag{2}$$

and note that $e = m$ when $c_{ij} = 0$ $(j \neq i)$, $e = 1$ when $c_{ij} = 1$, as expected. The adequacy of the chi-square approximation to the distribution of H will be examined later.

3. THE CORRELATION MATRIX c

What is the behaviour of **c** under increasing (absolute correlations in X?

This question was answered empirically, using Monte Carlo simulation. One thousand samples of correlated normal variate pairs were generated, for each of a set of 8 sample sizes (n = 10, 20, 50, 100, 200, 500, 1000 and 2000) and 19 parent correlations (ρ = ±0.995, ±0.99, ±0.95, ±0.9, ±0.8, ±0.7, ±0.6, ±0.4, ±0.2, 0.0). Wichmann and Hill's (1982) pseudo-random number generator provided uniformly distributed source variates. Values of z (normalized W) and hence k were calculated for each of the paired variates, and the sample correlation coefficient c between the k's evaluated. Thus, each sample size yielded 19 values of c, each based on 1000 k-pairs. In effect, the values of c for $\pm\rho$ are replicates, since c is unaffected by the sign of ρ.

A plot of c versus $|\rho|$ is given in Fig. 1 for $n = 100$. Evidently c is remarkably resistant to $|\rho|$ and n, and only begins to "take off" at about $|\rho| = 0.7$. Even at $|\rho| = 0.9$, c is only 0.4 at $n = 100$. This implies that for moderately correlated multivariate data, the marginal k (and W) values are virtually uncorrelated and may be considered separately as independent χ_1^2's, or summed to give the statistic $H \sim \chi_m^2$ as described above. This result simplifies both computation and interpretation.

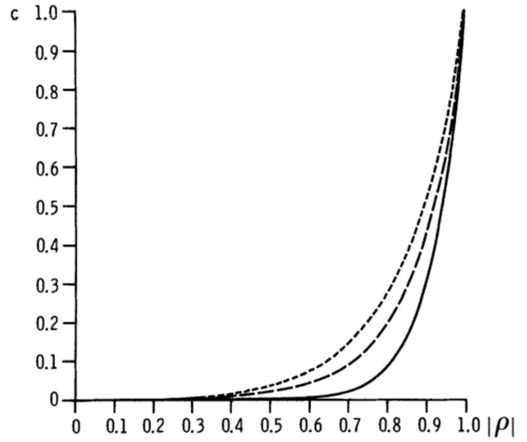

Fig. 1. The correlation (c) between pairs of transformed normalized W values is plotted against the absolute correlation ($|\rho|$) in the parent simulated bivariate normal distribution. Curve represents smoothed data for sample size $n = 100$.

4. APPROXIMATING THE RELATIONSHIP BETWEEN c AND $|\rho|$

For the purposes of a computer algorithm and in cases where $|\rho|$ is large, it is useful to have a specific function for calculating c from $|\rho|$ and n, enabling an estimated **c** matrix and H statistic (equation (2)) to be obtained from the sample correlation matrix. Thus (dropping the modulus sign for simplicity of notation)

$$c = g(\rho, n) \tag{3}$$

would be realized in practice as

$$\hat{c}_{ij} = \begin{Bmatrix} g(r_{ij}, n), & j \neq i \\ 1 & j = i \end{Bmatrix}, \quad i, j = 1, \ldots, m.$$

where $R = \{r_{ij}\}$ is the sample correlation matrix of X. The function g is subject to boundary conditions

$$g(0, n) = 0$$
$$g(1, n) = 1$$

for all n. A suitable class of models is

$$g(\rho, n) = \rho^{\lambda} \left[1 - \frac{\mu}{\nu} (1-\rho)^{\mu} \right], \quad \lambda, \mu, \nu > 0, \tag{4}$$

where λ, μ, and ν may be functions of n. Equation (4) was fitted to the simulated (c, ρ) pairs by maximum likelihood using *FIT MODEL* in the package MLP (Ross, 1980). Values of $\lambda = 5$ and $\mu = 0.715$ were found to be satisfactory throughout the chosen range of n (10 to 2000). Values of ν were smoothed as a cubic in $x = \log_e(n)$:

$$\nu(n) = 0.21364 + 0.015124\, x^2 - 0.0018034\, x^3$$

In fact, the (c, ρ) curves for different n are very similar and the compromise choice $\nu = 0.35$ gives a reasonable approximation throughout the range of n.

5. EMPIRICAL DISTRIBUTION OF H

The distribution of H for $m = 2$ was investigated. At each of three sample sizes $n = 20, 50, 100$ and 7 parent correlations $\rho = 0, 0.5, 0.75, 0.9, 0.95, 0.99, 0.995$, 1000 values of H were derived from simulated bivariate normal pairs. The empirical upper 2.5, 5 and 10 per cent points of H, based on each set of 1000 values, were determined and converted to equivalent p values for the appropriate χ^2_e distribution. These observed percentage points were usually smaller than the corresponding χ^2_e values; thus H is lighter in the tail than χ^2_e and gives rise to a conservative test. The results for $n = 20$, 50 and 100 were similar so were averaged for presentation in Fig. 2, where the abscissa is ρ, in the scale of Fisher's z transform $\frac{1}{2} \log((1+\rho)/(1-\rho))$ for clarity. The largest error occurs at about $\rho = 0.95$, where the H test gives $p = 0.12$, 0.067 and 0.037 compared with nominal values of 0.1, 0.05 and 0.025 respectively (Fig. 2).

Thus I consider the χ^2 approximation for H to be good enough for practical purposes. Further investigation with $m = 3$ confirmed this conclusion, but larger values of m have not yet been examined.

5.1. *The Geometry of H*

The definition of H as a test statistic implies the selection of a region, in the space of m-tuples $(z_1, \ldots, z_m)$ of normalized W statistics, for rejecting the hypothesis of m-normality of the parent data. The shape of this region for $m = 2$ is illustrated as contours of 5 per cent Type I error probability in Fig. 3, for $c = 0.0$ and 0.9, where $c = \text{corr}(k_1, k_2)$.

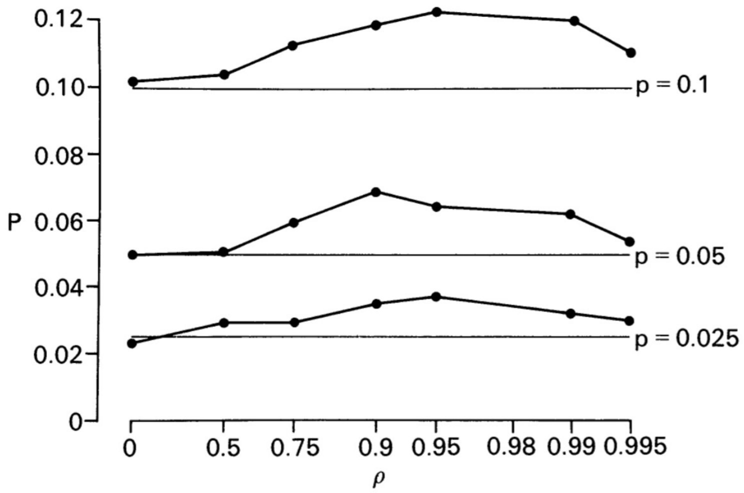

Fig. 2. Equivalent p values calculated on the assumption that H has a χ_e^2 distribution, based on the empirical upper 2.5, 5 and 10 per cent points of H. Abscissa is the absolute parent correlation $|\rho|$ in the scale of Fisher's z-transform. H has a lighter right-hand tail than χ_e^2.

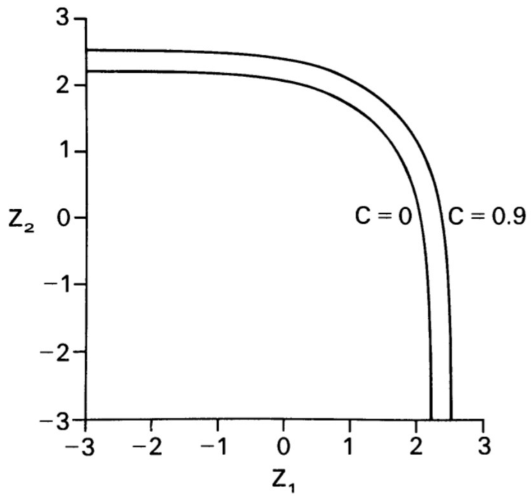

Fig. 3. Contours of constant H, representing $\alpha = 0.05$, in the (z_1, z_2) plane. The areas to the right of and above the curves form rejection regions at the 5 per cent level of the null hypothesis that the parent data were bivariate normally distributed. Left-hand curve, $c = 0.0$; right-hand curve $c = 0.9$ (see text).

Since W is a "one-tailed" test, the upper right-hand area in Fig. 3 represents one possible multivariate generalization of such one-sided tests. It in fact provides "guard values" for each variate in turn, with one arc (or curved surface for $m > 2$) connecting the relevant lines (or hyperplanes). Such tests will of course *not* be sensitive to observations that lie outside the conventional (2-sided) probability ellipse unless at least one of the variates has a large value. On the other hand, the test will be more sensitive to departures from the null hypothesis in the region around $z_1 = z_2$ than the equivalent two-sided version, as would be expected.

5.2. *Operational Example*

Suppose $m = 3, n = 77, R = \begin{pmatrix} 1 & 0.954 & 0.791 \\ 0.954 & 1 & 0.786 \\ 0.791 & 0.786 & 1 \end{pmatrix}$;

we wish to find $\mathbf{c}$ and e (equivalent degrees of freedom for H). For $n = 77$, $\log(n) = 4.344$,

$$\nu = 0.21364 + 0.01524(4.344)^2 - 0.0018034(4.344)^2$$
$$= 0.3512$$

$r_{12} = 0.954$, so from equation (4)

$$c_{12} = g(r_{12}, n) = g(0.954, 77)$$
$$= 0.954^5 [1 - 0.715 \times 0.954 \times (1 - 0.954)^{0.715} / 0.3512]$$
$$= 0.620.$$

Applying equation (4) to the other c_{ij},

$$\mathbf{c} = \begin{pmatrix} 1 & 0.620 & 0.147 \\ 0.620 & 1 & 0.141 \\ 0.147 & 0.141 & 1 \end{pmatrix}.$$

The mean of the c_{ij} for $i \neq j$ is $\bar{c} = 0.303$, hence

$$e = \frac{3}{1 + (3-1)\,0.303} = 1.868$$

If, for example, the normalized W values were (1.49, 2.36, 0.95), then $k = (3.328, 6.796, 1.874)$ and $H = (3.328 + 6.796 + 1.874) \times 1.868/3 = 7.471$, giving $p = 0.021$. Without adjusting for the correlations between the k_i, i.e. assuming $e = m = 3$, one gets $H = 12.00$ ($p = 0.007$); in this case, the size of the H-test is considerably affected by the magnitude of the r_{ij}.

6. TESTING SQUARED RADII

Healy (1968) pointed out that the quantities

$$r_i^2 = (x_i - \bar{x})' S^{-1} (x_i - \bar{x}), \; i = 1, \ldots, n,$$

where x_i is an m variate vector, $\bar{x}$ is the vector of variate means and S is the variance-covariance matrix of the x_i, are distributed approximately as χ_m^2 if the x_i are m-normal. He described normal probability plots of the (square-root transformed) r_i^2. Small (1978) showed that the distribution of the r_i^2 was actually proportional to a beta distribution, and provided an approximation for the expected order statistics thereof.

Let $z_i = \Phi^{-1}[F(r_i^2)]$, where F is a suitable CDF (χ^2 or beta) for r_i^2. If the x_i were m-normal, the z_i would be approximately independent $N(0, 1)$,.and could be normal plotted and subjected to the W test. I shall call this procedure the "Ω test" of m-normality for convenience, where Ω is the normal score corresponding to the above W value; $\Omega \sim N(0, 1)$ if the x_i were m-normal.

However, departures from m-normality may occur in lower dimensions of R^m and be masked in the plot of m-dimensional radii. This will be well-demonstrated in the examples to follow. A logical extension of the Ω test is to examine the plots of squared radii for subsets of p variates ($1 < p < m$). Each value of p gives rise to $K = K(m, p) = m!/[(m-p)!p!]$ distinct combinations of variates and the same number of (non-independent) Ω tests, $\Omega_1, \ldots, \Omega_K$, say. These scores may be inspected individually, or if there is a large number, they may be normal plotted, rather in the same spirit as the half-normal plot of large matrices of correlation coefficients (Hills, 1969). One may derive an appropriate χ^2 test statistic $\theta_p \sim \chi_K^2$ from the Ω_i in the same manner as equation (1), and with the same rationale:

$$\theta_p = \sum_{i=1}^{K} \{\Phi^{-1} [\tfrac{1}{2} \Phi(-\Omega_i)]\}^2$$

This technique should be regarded as exploratory, and it may give valuable insight into the multivariate structure of the data.

6.1. *Examples*

I shall now give three examples of the techniques in action.

Example 1. *Placental measurements and birth-weight*

Penfold *et al.* (1979) investigated the relationship between two measures of foetal well-being (standardized foetal birth-weight and Apgar score) and measurements of placental diameter, weight and umbilical cord length. Standardized birth-weight was calculated from observed weight corrected for gestational age, sex and maternal parity, height and body-weight, using the standards of Thomson *et al.* (1968). I shall examine six variates:

1. *PAT* = patient number (not analysed, of course, but see later).
2. *TWT* = trimmed placental weight.
3. *SIZE* = (length + width)$^{1/3}$.
4. *CLEN* = cord length.
5. *SDS* = standardized birth-weight.
6. *THIC* = (TWT)$^{1/3}$/SIZE (a measure of placental thickness).

A total of 491 cases were available. The correlation matrix for variates 2 to 6 is given in Table 1. None of the univariate W tests is remarkable, $H = 7.85$, $e = 4.36$ ($p = 0.12$), but the overall test of squared radii shows a highly significant departure from 5-normality, with $W = 0.9410$, $\Omega = 11.7$ ($p = 0$). Ω tests on all pairs of variates gave p-values of 0.61, 0.75, 0.028, 0.16, 0.40, 0.48, 0.056, 0.98 and 0.14, with $\theta_2 = 9.37$ ($p = 0.4$). However, analysis of triples showed one combination with $\Omega = 26.2$ ($p = 0$), that for variates 2, 3 and 6. One then notes that of course variate 6 (*THIC*) is a direct mathematical function of 2 and 3 (*TWT* and *SIZE*), and therefore the geometry of this configuration is far from random normal! This example shows the utility of examining subsets carefully. Excluding *THIC*, variates 2 to 5 show reasonable 4-normality; the test of squared radii gives $p = 0.041$; $H = 7.83$, $e = 3.91$ ($p = 0.093$) and the subset analyses are satisfactory.

TABLE 1
The correlation matrix for the placental measurements of Example 1. See text for key to the abbreviated variate-names

Variate	*TWT*	*SIZE*	*CLEN*	*SDS*	*THIC*
TWT	1				
SIZE	0.553	1			
CLEN	0.189	0.143	1		
SDS	0.548	0.463	0.068	1	
THIC	0.875	0.097	0.139	0.394	1

In the above case, the tests of squared radii were more informative than the combined univariate W tests. However, the reverse can easily be true; suppose variate 1 (*PAT*) had inadvertently been included in the analysis. *PAT* consisted simply of the integers 1 to 491. The Ω test of radii for variates 1 to 5 gives $\Omega = 1.56$ ($p = 0.06$)! Of course, W for variate 1 is very highly significant, but the overall radii test (and, incidentally, each of the tests on subsets) is insensitive to this departure from normality.

Example 2. *Some haematology data*

The data in this example form a small subset of a health survey of paint sprayers in a car assembly plant (Chatterjee *et al.* 1982).

Six variates are considered:
1. *HAEMO* = haemoglobin concentration.
2. *PCV* = packed cell volume.
3. *WBC* = white blood cell count.
4. *LYMPHO* = lymphocyte count.
5. *NEUTRO* = neutrophil count.
6. *LEAD* = serum lead concentration.

These were measured on 103 black (West Indian or African) workers. The data are given in full in Table 2. All correlations between the variates were less than 0.4 in absolute magnitude, except for $r_{12} = 0.78$, $r_{34} = 0.78$ and $r_{35} = 0.55$. Variates 3, 4, 5 and 6 all showed a skew distribution and were logarithmically transformed before analysis, as is common with haematological data.

TABLE 2
The full haematology data used in Example 2. Cases 21, 47 and 52 turned out to be outliers in R^3 (see text) on WBC, LYMPHO and NEUTRO

Case no.	*HAEMO*	*PCV*	*WBC*	*LYMPHO*	*NEUTRO*	*LEAD*
1	13.4	39	4100	14	25	17
2	14.6	46	5000	15	30	20
3	13.5	42	4500	19	21	18
4	15.0	46	4600	23	16	18
5	14.6	44	5100	17	31	19
6	14.0	44	4900	20	24	19
7	16.4	49	4300	21	17	18
8	14.8	44	4400	16	26	29
9	15.2	46	4100	27	13	27
10	15.5	48	8400	34	42	36
11	15.2	47	5600	26	27	22
12	16.9	50	5100	28	17	23
13	14.8	44	4700	24	20	23
14	16.2	45	5600	26	25	19
15	14.7	43	4000	23	13	17
16	14.7	42	3400	9	22	13
17	16.5	45	5400	18	32	17
18	15.4	45	6900	28	36	24
19	15.1	45	4600	17	29	17
20	14.2	46	4200	14	25	28
21	15.9	46	5200	8	34	16
22	16.0	47	4700	25	14	18
23	17.4	50	8600	37	39	17
24	14.3	43	5500	20	31	19
25	14.8	44	4200	15	24	19
26	14.9	43	4300	9	32	17
27	15.5	45	5200	16	30	20
28	14.5	43	3900	18	18	25
29	14.4	45	6000	17	37	23
30	14.6	44	4700	23	21	27
31	15.3	45	7900	43	23	23
32	14.9	45	3400	17	15	24
33	15.8	47	6000	23	32	21
34	14.4	44	7700	31	39	23
35	14.7	46	3700	11	23	23
36	14.8	43	5200	25	19	22
37	15.4	45	6000	30	25	18
38	16.2	50	8100	32	38	18
39	15.0	45	4900	17	26	24
40	15.1	47	6000	22	33	16
41	16.0	46	4600	20	22	22
42	15.3	48	5500	20	23	23
43	14.5	41	6200	20	36	21
44	14.2	41	4900	26	20	20
45	15.0	45	7200	40	25	25

TABLE 2 (continued)

Case no.	*HAEMO*	*PCV*	*WBC*	*LYMPHO*	*NEUTRO*	*LEAD*
46	14.2	46	5800	22	31	22
47	14.9	45	8400	61	17	17
48	16.2	48	3100	12	15	18
49	14.5	45	4000	20	18	20
50	16.4	49	6900	35	22	24
51	14.7	44	7800	38	34	16
52	17.0	52	6300	19	21	16
53	15.4	47	3400	12	19	18
54	13.8	40	4500	19	23	21
55	16.1	47	4600	17	28	20
56	14.6	45	4700	23	22	27
57	15.0	44	5800	14	39	21
58	16.2	47	4100	16	24	18
59	17.0	51	5700	26	29	20
60	14.0	44	4100	16	24	18
61	15.4	46	6200	32	25	16
62	15.6	46	4700	28	16	16
63	15.8	48	4500	24	20	23
64	13.2	38	5300	16	26	20
65	14.9	47	5000	22	25	15
66	14.9	47	3900	15	19	16
67	14.0	45	5200	23	25	17
68	16.1	47	4300	19	22	22
69	14.7	46	6800	35	25	18
70	14.8	45	8900	47	36	17
71	17.0	51	6300	42	19	15
72	15.2	45	4600	21	22	18
73	15.2	43	5600	25	28	17
74	13.8	41	6300	25	27	15
75	14.8	43	6400	36	24	18
76	16.1	47	5200	18	28	25
77	15.0	43	6300	22	34	17
78	16.2	46	6000	25	25	24
79	14.8	44	3900	9	25	14
80	17.2	44	4100	12	27	18
81	17.2	48	5000	25	19	25
82	14.6	43	5500	22	31	19
83	14.4	44	4300	20	20	15
84	15.4	48	5700	29	26	24
85	16.0	52	4100	21	15	22
86	15.0	45	5000	27	18	20
87	14.8	44	5700	29	23	23
88	15.4	43	3300	10	20	19
89	16.0	47	6100	32	23	26
90	14.8	43	5100	18	31	19
91	13.8	41	8100	52	24	17
92	14.7	43	5200	24	24	17
93	14.6	44	9899	69	28	18
94	13.6	42	6100	24	30	15
95	14.5	44	4800	14	29	15
96	14.3	39	5000	25	20	19
97	15.3	45	4000	19	19	16
98	16.4	49	6000	34	22	17
99	14.8	44	4500	22	18	25
100	16.6	48	4700	17	27	20
101	16.0	49	7000	36	28	18
102	15.5	46	6600	30	33	13
103	14.3	46	5700	26	20	21

The test of combined W values gave $H = 10.80$, $e = 5.46$ ($p = 0.08$), and the normal scores vector z was (2.02, 0.79, 0.63, −0.58, 0.61, 0.86). The six normal plots look reasonably linear. However, the transformed squared radii were strongly non-normal ($\Omega = 3.37$, $p = 0.0004$). What

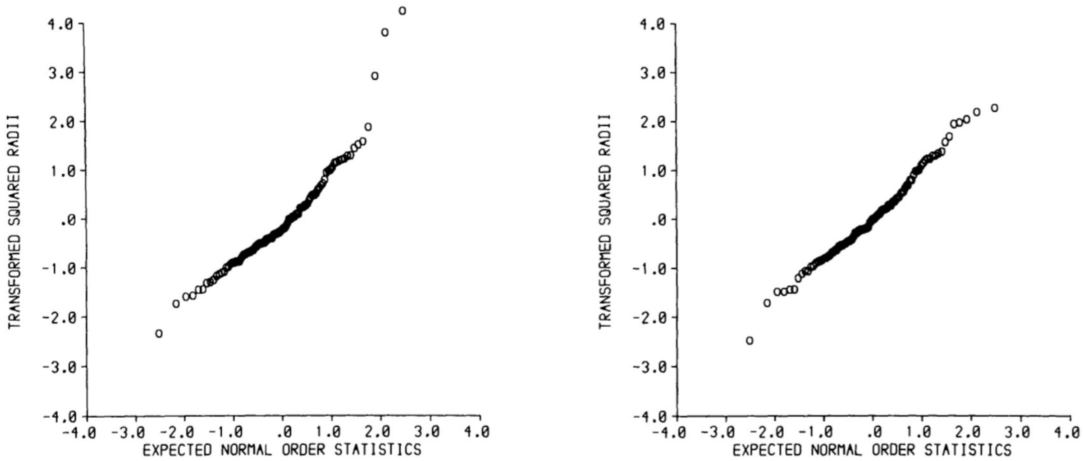

Fig. 4. Normal plot of square-root transformed radii for the haematology data of Example 2, variates 3, 4 and 5, (a) including 3 outliers (n = 103), (b) excluding 3 outliers (n = 100).

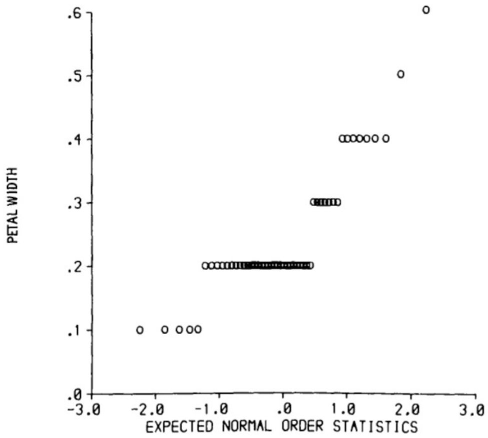

Fig. 5. Normal plot of variate 4 (petal width) for the *Iris setosa* data of Example 3. n = 50. Note heavy data-grouping.

was the source of departure from 6-normality? It could not apparently be attributed to the marginal distributions.

Significance levels for the Ω test on all pairs and triples of the variates are given in Table 3. The pairs show rather little ($\theta_2 = 17.55$, $p = 0.29$), but the triples display an astonishingly significant result for the squared radii from variates 3 (*WBC*), 4 (*LYMPHO*) and 5 (*NEUTRO*): $\Omega_{345} = 4.37$, $p = 6 \times 10^{-6}$. The associated normal plot of transformed radii (using the χ^2 approximation with 3 degrees of freedom for each radius) is given in Fig. 4(a). There are three pronounced outliers, which correspond in fact to subjects 21, 47 and 52 in Table 2. Fig. 4(b) shows the same plot as Fig. 4(a) after removing these three individuals and recalculating the radii. Now $\Omega_{345} = -0.05$ ($p = 0.52$), and the plot is satisfactory. Likewise, the Ω tests on the pairs, triples etc. give no evidence of departure from multivariate normality, for example, $\theta_3 = 7.94$ ($p = 0.99$), and the H test gave $p = 0.17$ with the 6 variates.

Example 3. *Iris setosa*

No paper on multivariate analysis would be complete without this famous data set, first used by Fisher (1936). The data were re-examined by Small (1980) who found that his indices (Q_1 and Q_2) of marginal skewness and kurtosis suggested significant departures for 4-normality at

TABLE 3

Results of Ω tests on squared radii for pairs and triples of haematological variates in Example 2. Values in the table are significance levels. n = 103. See text for key to variate numbers

Pairs					
Variate no.	2	3	4	5	6
1	0.85	0.32	0.65	0.081	0.72
2		0.13	0.052	0.12	0.50
3			0.24	0.995	0.55
4				0.25	0.34
5					0.63

Triples			
Variate nos.	*p*	*Variate nos.*	*p*
1, 2, 3	0.93	2, 3, 4	0.49
4	0.80	5	0.93
5	0.71	6	0.92
6	0.97	2, 4, 5	0.10
1, 3, 4	0.94	6	0.51
5	0.24	2, 5, 6	0.998
6	0.98	3, 4, 5	0.0000062
1, 4, 5	0.56	6	0.55
6	0.15	3, 5, 6	0.55
1, 5, 6	0.90	4, 5, 6	0.73

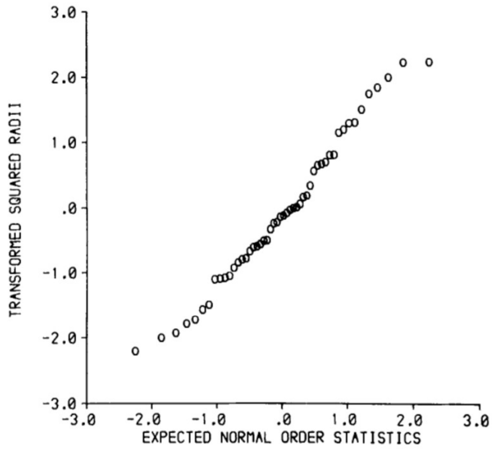

Fig. 6. Normal plot of square-root transformed radii for the four *Iris setosa* variates ($n = 50$).

$p = 0.027$ and $p = 0.072$ respectively. I consider *Iris setosa* only, as the data for the other two variates (*versicolor* and *virginica*) are less interesting.

Four variates (sepal length and width, petal length and width) were measured on $n = 50$ plants. The full data are given by Kendall (1975). The vector z (normalized scores) is (0.257, −0.264, 1.242, 5.581) for which $H = 34.4$, $e = 3.82$ ($p \ll 0.001$). Evidently, variate 4 (petal width) is markedly non-normal, and a normal plot (Fig. 5) shows both skewness and heavy grouping (there were only 6 distinct values, 0.1 – 0.6). W is known to be affected by grouping of the data (Pearson, D'Agostino and Bowman, 1977). Variates 1 to 3 show no peculiarities: $H = 3.37$, $e = 2.83$ ($p = 0.31$). Clearly variate 4 is the source of non-normality.

Despite variate 4, the plot of transformed squared radii (Fig. 6) is quite unremarkable, and the W test gives $\Omega = 0.61$ ($p = 0.27$). The p values for the Ω test on all pairs and triples of variates are given in Table 4. Any pair involving variate 4 indicates non-normality, whereas of the triples only Ω_{234} gives any suspicion. Again, variate 4 is implicated, but its effect becomes progressively diluted as the number of variates in the squared radii is increased.

TABLE 4
Results of Ω tests on squared radii for pairs and triples of iris variates, Example 3. Values are significance levels

Pairs

Variate no.	2	3	4
1	0.55	0.092	0.0060
2		0.11	0.00031
3			0.00027

Triples

Variate nos.	*p*
1, 2, 3	0.55
1, 2, 4	0.43
1, 3, 4	0.25
2, 3, 4	0.029

Key to variates: 1. Sepal length. 2. Sepal width. 3. Petal length. 4. Petal width.

7. COMMENT

The conclusion to be drawn from the examples is that departures from multivariate normality may be manifest only among subsets of variates, and therefore that combinations of variates should be examined separately. When many variates are present, the screening process ought to include all the univariate normal plots, the statistics H (combined marginal W tests), Ω (overall tests of squared radii) and as many of the subset Ω tests as are sensible; certainly θ_2 and θ_3 are desirable, but they should be accompanied by the corresponding normal plots of the Ω_{ij} and Ω_{ijk} values, which may display peculiarities to which the χ^2 tests (θ_2, θ_3) are insensitive. Since the Ω_{ij} etc. statistics are correlated, their individual significance levels and those of the θ statistics must be viewed as a rough guide only, to direct more detailed attention to particular aspects of the raw data.

A computer algorithm to calculate the H, Ω and θ statistics from an array of raw data is in preparation.

ACKNOWLEDGEMENT

The author thanks Dr D. S. Chatterjee, Ford Motor Co Ltd, Dagenham, UK, for supplying the haematological data used in Example 2, Dr F. E. Hytten for permission to use the placenta data of Example 1, and Professor M. J. R. Healy and a referee for useful comments on an earlier draft.

REFERENCES

Campbell, N. A. (1980) Robust procedures in multivariate analysis. I. Robust covariance estimation. *Appl. Statist.*, **29**, 231–237.

Chatterjee, D. S., Royston, J. P. and Smith, H. (1982) A health survey of paint sprayers in a car assembly plant. *Brit. J. Ind. Med.*, Submitted.

Cox, D. R. and Small, N. J. H. (1978) Testing multivariate normality. *Biometrika*, **65**, 263–272.

Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Ann. Eugenics*, VII, 179–188.

Healy, M. J. R. (1968) Multivariate normal plotting. *Appl. Statist.*, **17**, 157–161.

Hills, M. (1969) On looking at large correlation matrices. *Biometrika*, **56**, 249–253.

Kendall, M. G. (1975) *Multivariate Analysis*. London: Griffin.

Malkovich, J. F. and Afifi, A. A. (1973) On tests for multivariate normality. *J. Amer. Statist. Ass.*, **68**, 176–179.

Pearson, E. S., D'Agostino, R. B. and Bowman, K. O. (1977) Tests for departure from normality: comparison of powers. *Biometrika*, **64**, 231–246.

Penfold, P., Drury, L., Lewis, L., Royston, J. P. and Hytten, F. E. (1979) Case note descriptions of the placenta: are they worthwhile? *Br. J. Obstet. Gynaecol.*, **86**, 337–339.

Ross, G. J. S. (1980) *MLP. Maximum Likelihood Program*. Harpenden: Rothamsted Experimental Station. Ch. 10.

Royston, J. P. (1982) An extension of Shapiro and Wilk's *W* test for normality to large samples. *Appl. Statist.*, **31**, 115–124.

Sarhan, A. E. and Greenberg, B. G. (1956). Estimation of location and scale parameters by order statistics from singly and double censored samples. *Ann. Math. Statist.*, **27**, 427–451.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, **65**, 591–611.

Small, N. J. H. (1978) Plotting squared radii. *Biometrika*, **65**, 657–658.

———(1980) Marginal skewness and kurtosis in testing multivariate normality. *Appl. Statist.*, **29**, 85–87.

Wichmann, B. A. and Hill, I. D. (1982) An efficient and portable pseudo-random number generator. *Appl. Statist.*, **31**, 188–90.