

DOCUMENT RESUME

ED 067 406

TM 001 801

AUTHOR Aitkin, Murray A.  
TITLE A Class of Tests for Multivariate Normality Based on  
Linear Functions of Order Statistics.  
INSTITUTION Educational Testing Service, Princeton, N.J.  
REPORT NO ETS-RB-72-16  
PUB DATE Apr 72  
NOTE 22p.

EDRS PRICE MF-\$0.65 HC-\$3.29  
DESCRIPTORS \*Comparative Analysis; \*Correlation; Data Analysis;  
\*Mathematical Models; \*Statistical Analysis;  
Statistics; \*Tests  
IDENTIFIERS \*Normality Tests

ABSTRACT

A class of tests for normality using the ratio of two estimates of the standard deviation is generalized to provide a class of tests for multivariate normality using a characterization of the multivariate normal. The powers of some of the tests are examined numerically and compared with the power of a recent similar test.  
(Author)

ED 067406

**RESEARCH**

**Z-1111111**

TM 001 801

U.S. DEPARTMENT OF HEALTH,  
EDUCATION & WELFARE  
OFFICE OF EDUCATION  
THIS DOCUMENT HAS BEEN REPRO-  
DUCED EXACTLY AS RECEIVED FROM  
THE PERSON OR ORGANIZATION ORIG-  
INATING IT POINTS OF VIEW OR OPIN-  
IONS STATED DO NOT NECESSARILY  
REPRESENT OFFICIAL OFFICE OF EDU-  
CATION POSITION OR POLICY

PE-72-16

A CLASS OF TESTS FOR MULTIVARIATE NORMALITY BASED ON  
LINEAR FUNCTIONS OF ORDER STATISTICS

Murray A. Aitkin  
Macquarie University and Educational Testing Service

This Bulletin is a draft for interoffice circulation.  
Corrections and suggestions for revision are solicited.  
The Bulletin should not be cited as a reference without  
the specific permission of the author. It is automati-  
cally superseded upon formal publication of the material.

Educational Testing Service  
Princeton, New Jersey  
April 1972



A Class of Tests for Multivariate Normality Based on  
Linear Functions of Order Statistics

Murray A. Aitkin

Macquarie University and Educational Testing Service

SUMMARY

A class of tests for normality using the ratio of two estimates of the standard deviation is generalized to provide a class of tests for multivariate normality using a characterization of the multivariate normal. The powers of some of the tests are examined numerically and compared with the power of a recent similar test.

1. INTRODUCTION

Tests for normality based on the ratio of two estimates of the standard deviation from a single sample have been proposed by several authors. We restrict attention here to those statistics using the usual estimate based on the sum of squares in the denominator, and an estimate based on a linear function of the sample order statistics in the numerator. Such statistics include the ratio  $u$  of range to standard deviation (David, Hartley, and Pearson, 1954), Shapiro and Wilk's (1965) statistic  $W$  based on the best linear unbiased estimate of the standard deviation, and D'Agostino's (1971) statistic  $D$ , based on Gini's mean difference. Similar statistics could easily be constructed from other order statistic estimators of the standard deviation, for example, the mean deviation about the median (Nair, 1947).

Malkovich and Afifi (1971) have generalized the  $W$  statistic to the multivariate case using an approximation to a union-intersection test.

In this paper we generalize the class of statistics described above to the multivariate case by a union intersection procedure different from that considered by Malkovich and Afifi. Percentage points for some of these statistics are obtained by empirical sampling. The empirical powers of the tests are examined for certain nonnormal alternatives considered by Malkovich and Afifi.

## 2. TESTS FOR NORMALITY

Let  $X_{1n} \leq X_{2n} \leq \dots \leq X_{nn}$  be the order statistics in a sample of size  $n$  from a normal population  $N(\mu, \sigma^2)$ , and let  $b_{1n} \leq b_{2n} \leq \dots \leq b_{nn}$  be a set of constants. The sample "correlation" between the  $X_{in}$  and the  $b_{in}$  is

$$\frac{\sum_{i=1}^n (b_{in} - \bar{b})(X_{in} - \bar{X})}{\left\{ \sum_{i=1}^n (b_{in} - \bar{b})^2 \sum_{i=1}^n (X_{in} - \bar{X})^2 \right\}^{1/2}}$$

where

$$\bar{b} = \frac{1}{n} \sum_{i=1}^n b_{in} \quad , \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_{in} \quad .$$

Defining

$$a_{in} = (b_{in} - \bar{b}) / \left\{ \sum_1^n (b_{in} - \bar{b})^2 \right\}^{1/2} ,$$

and defining

$$\sum_1^n (X_{in} - \bar{X})^2 = s^2 ,$$

the sum of squared deviations about the mean, we may express the "correlation" as

$$r(\underline{a}, n) = \sum_1^n a_{in} X_{in} / s ,$$

where now

$$\sum_1^n a_{in} = 0 , \quad \sum_1^n a_{in}^2 = 1 .$$

Several test statistics proposed for testing normality may be put in the above form. David, Hartley, and Pearson (1954) considered

$$\begin{aligned} u &= (X_{nn} - X_{1n}) / \{s/(n-1)\}^{1/2} \\ &= \{2(n-1)\}^{1/2} r(\underline{a}_1, n) \end{aligned}$$

where  $\underline{a}'_1 = (-2^{-1/2}, 0, 0, \dots, 0, 2^{-1/2})$ . Shapiro and Wilk (1965) consider

$$W = r^2(a_{-2}, n) ,$$

where

$$a_{-2} = V^{-1} \underline{m} / \{\underline{m}' V^{-2} \underline{m}\}^{1/2} ,$$

$\underline{m}$  and  $V$  being, respectively, the mean vector and covariance matrix of the vector of normal order statistics. D'Agostino (1971) considers

$$D = \frac{\sum_{i=1}^n \{i - \frac{1}{2}(n+1)\} X_{in}}{n^{3/2} s}$$

$$= \{(n^2 - 1)/12n^2\}^{1/2} r(a_3, n)$$

where

$$a_{3in} = \{i - \frac{1}{2}(n+1)\} / \{(n(n^2 - 1)/12)\}^{1/2} .$$

Pearson and Chandra Sekar (1936) consider

$$\tau = n^{1/2} (X_{nn} - \bar{X}) / s$$

$$= (n - 1)^{1/2} r(a_4, n) ,$$

where

$$a_{4in} = \begin{cases} -\{n(n-1)\}^{-1/2} , & \text{for } i \neq n , \\ \{(n-1)/n\}^{1/2} , & \text{for } i = n . \end{cases}$$

Other such statistics can easily be constructed. For example, Nair (1947) considers  $m'$ , the mean deviation about the median, as an estimate of  $\sigma$ :

$$m' = \sum_{i=1}^n (X_{in} - \hat{\nu}) = \sum_{i=1}^k (X_{n-i+1,n} - X_{in}) ,$$

where  $\hat{\nu}$  is the sample median, and  $k = \frac{n-1}{2}$  if  $n$  is odd,  $k = \frac{n}{2}$  if  $n$  is even. A test for normality could then be based on

$$r(\underline{a}, n) = m' / (2k)^{1/2} s$$

### 3. TESTS FOR MULTIVARIATE NORMALITY

Let  $\underline{X}$  be a  $p$ -component random vector. It is well known (see, for example, Anderson, 1958) that  $\underline{X}$  is multivariate normal if and only if every linear function  $\underline{\ell}'\underline{X}$  is univariate normal, where  $\underline{\ell}$  is an arbitrary fixed vector. We use this property to construct a union-intersection test for normality.

Let  $X_1, \dots, X_n$  be a random sample from a population. Let  $\underline{\ell}$  be an arbitrary fixed vector, and define

$$Z^*_i = \underline{\ell}'X_i , \quad i = 1, \dots, n ,$$

the dependence of  $Z^*$  on  $\underline{\ell}$  being understood. Let the ordered  $Z^*_i$  be denoted by  $Z^*_{1n} \leq \dots \leq Z^*_{nn}$ . We construct any of the statistics

$$\sum_{i=1}^n a_{in} Z^*_{in} / \left\{ \sum_{i=1}^n (Z^*_{in} - \bar{Z}^*)^2 \right\}^{1/2}$$

$$= \sum_{i=1}^n a_{in} Z^*_{in} / \{ \underline{\ell}'S\underline{\ell} \}^{1/2} ,$$

where  $S$  is the matrix of sample cross-products. It is clear that such statistics are origin and scale invariant for all  $\underline{L}$ , i.e., do not depend on  $\underline{L}'\underline{\mu}$  and  $\underline{L}'\Sigma\underline{L}$  for all  $\underline{L}$ , where  $\underline{\mu}$  and  $\Sigma$  are the mean and covariance matrix of  $\underline{X}$ . Hence they do not depend on  $\underline{\mu}$  and  $\Sigma$ . Without changing the problem, we may therefore make a linear transformation

$$\underline{Y} = A\underline{X} + \underline{b} \quad ,$$

where  $A$  is a  $p \times p$  positive definite matrix and  $\underline{b}$  an arbitrary vector. Take  $A = S^{-1/2}$  and  $\underline{b} = S^{-1/2}\bar{\underline{X}}$ , where  $S^{1/2}$  is the (unique) symmetric square root of  $S$ , so that

$$\underline{Y}_i = S^{-1/2}(\underline{X}_i - \bar{\underline{X}}) \quad .$$

The components of  $\underline{Y}_i$  are then scores on standardized principal component variables with  $\bar{\underline{Y}} = \underline{0}$ ,  $\sum_i (\underline{Y}_i - \bar{\underline{Y}})(\underline{Y}_i - \bar{\underline{Y}})' = I$ . Define

$$\underline{Z}_i = \underline{L}'\underline{Y}_i \quad i = 1, \dots, n$$

and denote the ordered  $Z_i$  by  $Z_{1n} \leq \dots \leq Z_{nn}$ . Then we consider the statistics

$$r(\underline{a}, \underline{L}, n) = \sum_1^n a_{in} Z_{in} / (\underline{L}'\underline{L})^{1/2} \quad .$$

For different choices of  $\underline{a}$ , a one-sided or two-sided test may be appropriate (for example, the test based on  $W$  is one-sided, but that based on  $u$  two-sided). If small values of  $r(\underline{a}, \underline{L}, n)$  are significant, we do not reject a level  $\alpha$  the hypothesis of normality of  $\underline{L}'\underline{X}$  if

$$r_L^{1-\alpha}(\underline{a}, n) < r(\underline{a}, \underline{L}, n) \quad .$$

A size  $\alpha$  union-intersection test for the hypothesis of multivariate normality of  $\underline{X}$  will not reject if

$$c_L^{1-\alpha}(\underline{a}, n) < r(\underline{a}, \underline{L}, n)$$

for all  $\underline{L}$ , so that

$$\inf_{\underline{L}} r(\underline{a}, \underline{L}, n) > c_L^{1-\alpha}(\underline{a}, n) \quad ,$$

where  $c_L^{1-\alpha}(\underline{a}, n)$  is the lower  $100\alpha$  per cent point of the distribution of  $\inf_{\underline{L}} r(\underline{a}, \underline{L}, n)$ . A conservative two-sided test at level  $\alpha$  may be obtained by not rejecting the normality hypothesis when

$$\inf_{\underline{L}} r(\underline{a}, \underline{L}, n) > c_L^{1-\alpha/2}(\underline{a}, n) \quad ,$$

$$\sup_{\underline{L}} r(\underline{a}, \underline{L}, n) < c_U^{\alpha/2}(\underline{a}, n) \quad ,$$

where  $c_U^{\alpha}(\underline{a}, n)$  is the upper  $100\alpha$  per cent point of the distribution of  $\sup_{\underline{L}} r(\underline{a}, \underline{L}, n)$ .

#### 4. SUP r AND INF r FOR $p = 2$

Consider first the case  $p = 2$ . The linear function  $Z = \underline{l}'\underline{Y}$  is then  $l_1 Y_1 + l_2 Y_2$ . Since  $r(\underline{a}, \underline{L}, n)$  is independent of scale, we may assume without any loss of generality that  $l_1 = 1$ ,  $l_2 = l$ , since it is only the ratio  $l_2/l_1$  that matters. (We might instead set  $l_1 = \sin \theta$ ,  $l_2 = \cos \theta$  and map each point into a trigonometric function. Such a procedure has been used by Andrews (1971) in another context. The results are equivalent for  $p = 2$ , but for  $p > 2$  the polar transformation becomes less convenient.) We consider therefore the values

$$Z_1 = Y_{11} + \lambda Y_{21}, \dots, Z_n = Y_{1n} + \lambda Y_{2n} .$$

These may be plotted as  $n$  lines in the  $(Z, \lambda)$  plane. The  $n$  lines intersect in  $N = \binom{n}{2}$  points, defining  $N$  values of  $\lambda$ ,  $\lambda_1 \leq \dots \leq \lambda_N$  (some of which may be coincident). The points define  $N + 1$  regions  $L_j = \{\lambda : \lambda_{j-1} < \lambda \leq \lambda_j\}$ ,  $j = 1, \dots, N + 1$ , where  $\lambda_0 = -\infty$ ,  $\lambda_{N+1} = \infty$ . The regions  $L_j$  have the property that the ordering of the  $Z_i = \lambda Y_{-i}$  is the same for all  $\lambda \in L_j$ .

Now in the region  $L_j$ , let the ordered  $Z_i$  be

$$Z_{1n}^j \leq \dots \leq Z_{nn}^j ,$$

and write

$$Z_{in}^j = Y_{li}^j + \lambda Y_{2i}^j .$$

Then for  $\lambda \in L_j$ ,

$$\begin{aligned} r(\underline{a}, \lambda, n) &= r(\underline{a}, \lambda, n) = \left( \sum_1^n a_{in} Y_{li}^j + \lambda \sum_1^n a_{in} Y_{2i}^j \right) / (1 + \lambda^2)^{1/2} \\ &= (S_1^j + \lambda S_2^j) / (1 + \lambda^2)^{1/2} , \end{aligned}$$

where

$$S_k^j = \sum_1^n a_{in} Y_{ki}^j , \quad k = 1, 2 .$$

Now

$$\inf_{\lambda} r(\underline{a}, \lambda, n) = \inf_j \inf_{\lambda \in L_j} r(\underline{a}, \lambda, n) ,$$

and for  $\lambda \in L_j$ ,  $r(\underline{a}, \lambda, n)$  is the ratio of a linear form to the square root of a positive definite quadratic form. It is well-known that this ratio, for unrestricted  $\lambda$ , has only one extreme value, a maximum equal to  $(s_1^{j2} + s_2^{j2})^{1/2}$ , which occurs when  $\lambda = s_2^j/s_1^j$ . Hence

$$\inf_{\lambda} r(\underline{a}, \lambda, n) = \inf_j r(\underline{a}, \lambda_j, n) .$$

If  $\lambda = s_2^j/s_1^j \notin L_j$ , the supremum of  $r(\underline{a}, \lambda, n)$  will also occur at an  $\lambda_j$ . If  $\lambda = s_2^j/s_1^j \in L_j$ , write  $\lambda_j^* = s_2^j/s_1^j$ , and then

$$\sup_{\lambda} r(\underline{a}, \lambda, n) = \sup_j \{r(\underline{a}, \lambda_j, n), r(\underline{a}, \lambda_j^*, n)\} .$$

REMARK

The supremum simplifies considerably for the  $u$  statistic of David, Hartley and Pearson and for the  $\tau$  statistic of Pearson and Chandra Sekar. We have

$$\sup_{\lambda} r_u = \sup_{\lambda} r(\underline{a}_1, \lambda, n) = 2^{-1/2} \sup_{\lambda} (Z_{nn} - Z_{1n}) / ((\lambda' \lambda)^{1/2}) ,$$

and if we use the normalization  $\lambda' \lambda = 1$ ,

$$\sup_{\lambda} r_u = 2^{-1/2} \sup_{\lambda} (Z_{nn} - Z_{1n}) .$$

Thus  $\sup_{\lambda} r_u$  is a multiple of the greatest distance between the projections of any two points  $Y_{-1}, Y_j$ , on the hyperplane  $Z = \lambda' Y$ . The distance will be greatest when the hyperplane passes through the two points which are

farthest apart, i.e., at the diameter of the convex hull of the points  $\underline{y}_1, \dots, \underline{y}_n$ . Thus

$$\begin{aligned} \sup_{\underline{L}} r_u &= 2^{-1/2} \sup_{i,j} \{(\underline{y}_i - \underline{y}_j)'(\underline{y}_i - \underline{y}_j)\}^{1/2} \\ &= 2^{-1/2} \sup_{i,j} \{(\underline{x}_i - \underline{x}_j)'S^{-1}(\underline{x}_i - \underline{x}_j)\}^{1/2} \end{aligned}$$

in terms of the original variables.

Similarly,

$$\begin{aligned} \sup_{\underline{L}} r_\tau &= \sup_{\underline{L}} r(\underline{a}_t, \underline{L}, n) \\ &= (1 - \frac{1}{n})^{-1/2} \sup_{\underline{L}} \{Z_{nn}/(\underline{L}'\underline{L})^{1/2}\} \end{aligned}$$

and taking  $\underline{L}'\underline{L} = 1$  as above,

$$\sup_{\underline{L}} r_\tau = (1 - \frac{1}{n})^{-1/2} \sup_{\underline{L}} Z_{nn}.$$

Thus  $\sup_{\underline{L}} r_\tau$  is a multiple of the greatest distance from the origin of the projection of any point on the hyperplane  $Z = \underline{L}'\underline{Y}$ . This distance will be greatest for the point furthest from the origin. Thus

$$\begin{aligned} \sup_{\underline{L}} r_\tau &= (1 - \frac{1}{n})^{-1/2} \sup_i \{(\underline{y}_i' \underline{y}_i)\}^{1/2} \\ &= (1 - \frac{1}{n})^{-1/2} \sup_i \{(\underline{x}_i - \bar{\underline{X}})'S^{-1}(\underline{x}_i - \bar{\underline{X}})\}^{1/2} \end{aligned}$$

in terms of the original variables. These results hold for all  $p$ , but no corresponding results hold for  $\inf_{\underline{L}} r_u$  or  $\inf_{\underline{L}} r_\tau$ , or for  $\sup_{\underline{L}} r(\underline{a}, \underline{L}, n)$  in general, since the other statistics do not have a simple distance interpretation. Thus for  $r_u$ , no explicit calculation of the supremum is

necessary, for it must occur for one of the vectors  $\lambda$  already required for the infimum. For  $r_\tau$ , the explicit calculation is necessary unless  $Y_i = \bar{Y}$  for some  $i$ .

### 5. AN EXAMPLE

Below appears part of Student's data on the number of hours increase in sleep gained from the use of two drugs (Anderson, 1958, p. 51). For simplicity we have taken only the first five subjects.

<u>Patient</u>	<u>Drug A (<math>X_1</math>)</u>	<u>Drug B (<math>X_2</math>)</u>	<u><math>Y_1</math></u>	<u><math>Y_2</math></u>
1	1.9	0.7	0.58592	0.50966
2	0.8	-1.6	0.23394	-0.68989
3	1.1	-0.2	0.18796	0.10817
4	0.1	-1.2	-0.33245	-0.31752
5	-0.1	-0.1	-0.67537	0.38958

The matrix of sample cross products  $S$  is

$$S = \begin{bmatrix} 2.592 & 1.544 \\ 1.544 & 3.388 \end{bmatrix}, \quad S^{-1/2} = \begin{bmatrix} 0.704,134 & -0.183,727 \\ -0.183,727 & 0.609,414 \end{bmatrix}$$

while  $\bar{X}' = (0.76, -0.48)$ . The principal component scores  $Y_i = S^{-1/2}(X_i - \bar{X})$  are given above in the columns  $Y_1, Y_2$ . The  $\binom{5}{2} = 10$  points of intersection  $\lambda_j$  are obtained by solving simultaneously the pairs of equations

$$Y_{1i} + \lambda Y_{2i} = Y_{1j} + \lambda Y_{2j},$$

giving

$$\lambda = -\frac{Y_{1i} - Y_{1j}}{Y_{2i} - Y_{2j}}, \quad i \neq j = 1, \dots, 5.$$

(If  $Y_{2i} - Y_{2j}$  should be zero,  $Z$  is defined by  $Z_i = Y_{2i}$ .) The 10 values of  $\lambda$ , ordered from smallest to largest, are

-10.50375	0.05761
-1.22251	0.48497
-1.11024	0.84237
-0.99121	1.52104
-0.29343	3.06787

Suppose we wish to find  $\sup_{\lambda} r_u$  and  $\inf_{\lambda} r_u$ . Taking the value  $\lambda_1 = -10.50375$ , corresponding to the intersection of  $Z_1$  and  $Z_5$ , the  $Z_i$  are  $-4.76742, 7.48037, -0.94823, 3.00270, -4.76742$ . Then  $Z_{55}^1 - Z_{15}^1 = 12.24779$ ,  $\{2(1 + \lambda^2)\}^{1/2} = 14.92171$ ,  $r(\underline{a}_1, \lambda_1, n) = 0.8208$ . On repeating the calculation for the other nine values of  $\lambda$ , the following  $r$  values are obtained:

.8208	.8953
.9979	.8395
.9975	.8735
.9940	.8455
.8319	.8836

To test for a local maximum in  $L_1$ , the ordering of the  $Z_i: Z_1 < Z_5 < Z_3 < Z_4 < Z_2$  defines the orderings

0.58592	0.50966
-0.67537	0.38958
0.18796	0.10817
-0.33245	-0.31752
0.23394	-0.68989

of  $Y_1$  and  $Y_2$  respectively. The corresponding values of  $S_1^1$  and  $S_2^1$  are  $2^{-1/2}(0.23394 - 0.58592) = -0.24889$  and  $2^{-1/2}(-0.68989 - 0.50966) = -0.84820$  respectively. Then  $\lambda = S_2^1/S_1^1 = 3.40793 \notin L_1$ . Hence there is no local maximum for  $\lambda \in L_1$ . On repeating this procedure for the other  $L_j$ , only one local maximum occurs, namely  $r = 0.8740$  when  $\lambda = 0.90070$  in  $L_9$ . This does not yield a global maximum, so that

$$\inf_{\lambda} r(\underline{a}_1, \lambda, n) = .8208 \text{ at } \lambda = -10.50375 ,$$

$$\sup_{\lambda} r(\underline{a}_1, \lambda, n) = .9979 \text{ at } \lambda = -1.22251 .$$

To verify the supremum in this case, we note that . . .

$$2^{-1/2} \{ (\underline{y}_2 - \underline{y}_5)' (\underline{y}_2 - \underline{y}_5) \}^{1/2} = .9980 ,$$

which agrees with the above result within round-off error.

## 6. GENERAL $p$

The argument in §4 can be extended directly to any  $p$  . Without loss of generality we consider the standardized principal variables  $\underline{y}_i = S^{-1/2} (\underline{x}_i - \bar{X})$  , and again take the linear function  $\lambda' \underline{y}$  as  $\lambda_1 Y_1 + \lambda_2 Y_2 + \dots + \lambda_{p-1} Y_{p-1}$  . The values  $Z_i = \lambda_1 Y_{1i} + \lambda_2 Y_{2i} + \dots + \lambda_{p-1} Y_{pi}$  now define  $n$  hyperplanes. These intersect  $p$  at a time in  $\binom{n}{p}$  points, defining  $\binom{n}{p}$  values of  $\lambda$  , which then define regions with the property that the ordering of the  $Z_i$  is fixed in each region. Again the infimum of  $r(\underline{a}, \lambda, n)$  will occur at one of the vertices of the region, and the supremum will occur either at one of these points, or at a local maximum if one occurs in a region, the maximum then being  $(\sum_{k=1}^p S_k^{j2})^{1/2}$  , occurring when  $\lambda_k = S_k^j / S_1^j$  , where  $S_k^j = \sum_{i=1}^n a_{in} Y_{ki}^j$  ,  $k = 1, \dots, p$  , the  $j$  superscript denoting the  $j$ -th region.

However, the ordering of the  $Z_i$  at each of the  $\binom{n}{p}$  values of  $\lambda$  becomes a major computational problem for  $p > 2$  . For  $p = 2$  , it is necessary to order the  $Z_i$  only once, for if the ordering in (say)  $L_0$  is

established, then it is known that in  $L_1$ , the ordering of just two observations is interchanged from that in  $L_0$ , and these observations are identified by the original determination of  $\lambda_1$ . Thus the ordering of the  $Z_i$  in every  $L_j$  can be obtained with just one ordering. This does not happen for  $p > 2$ , however, essentially because the points cannot be ordered on one dimension. The ordering must therefore be recalculated at each of the  $\binom{n}{p}$  values of  $\lambda$ , a very time-consuming process. Some saving in time is possible for  $r_u$ , for this requires for each  $\lambda$  only the maximum and minimum  $Z$ , not a complete ordering.

We therefore consider an alternative procedure. Let

$$Z = Y_1 + \lambda Y_2 + \lambda^2 Y_3 + \dots + \lambda^{p-1} Y_p .$$

Then  $\underline{Y}$  is mapped into a polynomial in the  $(Z, \lambda)$  plane. The results of §4 may now be extended with some changes, but we note that all possible linear functions  $\lambda' \underline{Y}$  cannot be generated in this way. Thus this procedure may be less powerful against certain kinds of departures from normality than one which considers all linear functions.

The  $n$  polynomials  $Z_i$  now intersect two at a time in  $p - 1$  points, defining  $(p - 1) \binom{n}{2}$  points  $\lambda_j$  (some of which may be coincident or imaginary, corresponding to complex roots in  $\lambda$  of  $Z_i - Z_j = 0$ ). The regions  $L_j$  are defined as before, and in  $L_j$ ,

$$r(\underline{a}, \lambda, n) = \left( \sum_{k=1}^p s_k^j \lambda^{k-1} \right) / \left( \sum_{k=1}^p \lambda^{2(k-1)} \right)^{1/2} .$$

Further complications now arise as this function has multiple maxima and minima. We shall ignore these however and consider only the points  $\lambda_j$  defined above. Again this may result in some loss of power: this question is examined in §8. We thus evaluate

$$\inf_{\lambda_j} r(\underline{a}, \lambda_j, n) \quad ,$$

and 
$$\sup_{\lambda_j} r(\underline{a}, \lambda_j, n) \quad ,$$

and accept or reject the hypothesis of multivariate normality accordingly.

### 7. OTHER TESTS FOR MULTIVARIATE NORMALITY

Tests for multivariate normality have recently been considered by Malkovich and Afifi (1971), in a study including the tests discussed in Kowalski (1970) and some others. In particular, they generalize the skewness and kurtosis statistics  $b_1^{1/2}$  and  $b_2$  to the multivariate case by a union-intersection argument. Shapiro and Wilk's  $W$  is also generalized, but by a procedure different from that described in §4 and §6. Rather than obtaining the infimum of  $W$  over all linear functions  $\lambda'X$ , they consider the linear function  $\lambda'X$  which produces a value of  $W$  as close to its lower bound as possible. It is known (see Shapiro and Wilk (1965) for details) that  $W$  attains its lower bound when

$$X_j - \bar{X} = (n - 1)/(na_{1n}) \quad ,$$

$$X_i - \bar{X} = -1/(na_{1n}) \quad ,$$

for  $i = 1, \dots, n$ ,  $i \neq j$ , for any  $j$ . Malkovich and Afifi consider the vector  $\lambda$  which minimizes

$$[\lambda'(X_j - \bar{X}) - (n - 1)/(na_{1n})]^2 + \sum_{i \neq j} [\lambda'(X_i - \bar{X}) + 1/(na_{1n})]^2 \quad ,$$

which is

$$\lambda = S^{-1}(X_j - \bar{X})/a_{1n} \quad .$$

Since  $j$  may take any value from 1 to  $n$ ,  $\underline{X}_j$  is chosen to maximize the denominator of  $W$ . Thus let  $m$  be such that

$$(\underline{X}_m - \bar{X})'S^{-1}(\underline{X}_m - \bar{X}) = \max_{1 \leq j \leq n} (\underline{X}_j - \bar{X})'S^{-1}(\underline{X}_j - \bar{X}) .$$

Let the ordered values of

$$U_j = (\underline{X}_j - \bar{X})'S^{-1}(\underline{X}_m - \bar{X})$$

be  $U_{1n} \leq \dots \leq U_{nn}$ . Then the Malkovich-Afifi statistic is

$$W_p = [\sum_{i=1}^n U_{in}]^2 / (\underline{X}_m - \bar{X})'S^{-1}(\underline{X}_m - \bar{X}) .$$

If we use the standardized principal variables

$$\underline{Y}_i = S^{-1/2}(\underline{X}_i - \bar{X}) ,$$

this reduces to

$$W_p = [\sum_{i=1}^n V_{in}]^2 / \underline{Y}_m' \underline{Y}_m ,$$

where

$$V_j = \underline{Y}_j' \underline{Y}_m ,$$

$V_{1n} \leq \dots \leq V_{nn}$  are the ordered  $V_j$ , and  $m$  is such that

$$\underline{Y}_m' \underline{Y}_m = \max_{1 \leq j \leq n} \underline{Y}_j' \underline{Y}_j .$$

The null distributions of the statistics  $r$  described in §4 and §6 seem analytically intractable, as the corresponding univariate statistics in general do not have simple forms, and the multivariate statistics are obtained by data-based linear functions of  $\underline{X}$ . Percentage points of the

statistics were therefore obtained by simulation, using samples of 500. A more accurate table of percentage points is in preparation, but the results from samples of 500 should give a clear picture of power properties.

### 8. POWER RESULTS

Approximate lower percentage points of Malkovich and Afifi's  $W_p$ , of  $W_{\min} = \inf_{\underline{\ell}} r^2(a_2, \underline{\ell}, n)$ , and of  $u_{\min} = \{2(n-1)\}^{1/2} \inf_{\underline{\ell}} r(a_1, \underline{\ell}, n)$ , and approximate upper percentage points of  $u_{\max} = \{2(n-1)\}^{1/2} \sup_{\underline{\ell}} r(a_1, \underline{\ell}, n)$  were determined by generating 500 samples for  $p = 2$ ,  $n = 10, 20$ , and  $p = 3$ ,  $n = 10$ , at values of  $\alpha = .01, .02, .025, .05(.05), .25$ . In addition, approximate percentage points of  $W_{\min}^*$ ,  $u_{\min}^*$ , and  $u_{\max}^*$  were determined for  $p = 3$ ,  $n = 10$  and the same values of  $\alpha$ , where the asterisk indicates the use of a polynomial mapping rather than a hyperplane mapping of  $X$ .

The powers of the above tests were then determined against the following distributions.

$p = 2$ :	LN	$X_i$ independent log normal
	U	$X_i$ independent uniform
	$t_4$	$X_i$ independent $t_4$
	$t_{10}$	$X_i$ independent $t_{10}$
	N - LN	$X_1$ normal, $X_2$ log normal, $X_1, X_2$ independent
	N - U	$X_1$ normal, $X_2$ uniform, $X_1, X_2$ independent
	N - $t_4$	$X_1$ normal, $X_2$ $t_4$ , $X_1, X_2$ independent
	N - $t_{10}$	$X_1$ normal, $X_2$ $t_{10}$ , $X_1, X_2$ independent
	BVN(.5, .5)	$(X_1, X_2)$ mixture of bivariate normals where $BVN(p, \rho)$ has density
	BVN(.75, .5)	

$$p f_{\rho}(x_1, x_2) + (1 - p) f_{-\rho}(x_1, x_2)$$

where

$$f_{\rho}(x_1, x_2) = (2\pi)^{-1} (1 - \rho^2)^{-1/2} \exp\{- (x_1^2 - 2\rho x_1 x_2 + x_2^2) / 2(1 - \rho^2)\}$$

$p = \beta$  : LN , U ,  $t_{10}$  .

In Table 1 appear the results for  $\alpha = .10$  ,  $p = 2$  ,  $n = 10$  and  $20$  .

Other values of  $\alpha$  gave comparable results

TABLE 1  
EMPIRICAL POWERS FOR  $\alpha = .10$  ,  $p = 2$  ,  $n = 10$  AND  $20$

n = 10					
	<u>W<sub>p</sub></u>	<u>W<sub>min</sub></u>	<u>u<sub>min</sub>, u<sub>max</sub></u>	<u>W</u>	<u>u</u>
LN	.75	.78	.21	.72	.13
U	.08	.12	.18	.15	.32
$t_4$	.30	.26	.17	.23	.17
$t_{10}$	.18	.14	.13	.13	.13
N-LN	.53	.54	.13		
N-U	.13	.14	.19		
N- $t_4$	.20	.16	.14		
N- $t_{10}$	.15	.16	.15		
BVN(.5, .5)	.15	.13	.13		
BVN(.75, .5)	.14	.12	.12		
n = 20					
	<u>W<sub>p</sub></u>	<u>W<sub>min</sub></u>	<u>u<sub>min</sub>, u<sub>max</sub></u>	<u>W</u>	<u>u</u>
LN	.97	.99	.47	.96	.14
U	.03	.22	.51	.39	.53
$t_4$	.38	.36	.33	.27	.24
$t_{10}$	.30	.25	.21	.18	.14
N-LN	.81	.87	.22		
N-U	.09	.18	.34		
N- $t_4$	.26	.24	.21		
N- $t_{10}$	.23	.22	.21		
BVN(.5, .5)	.17	.16	.18		
BVN(.75, .5)	.16	.14	.15		

The columns headed  $W$  and  $u$  give the empirical powers for  $n = 10$  and  $20$  in the univariate case for  $W$  and  $u$ , reproduced from Shapiro, Wilk and Chen (1968). The results for  $\alpha = .10$ ,  $p = 3$ ,  $n = 10$  are given in Table 2.

TABLE 2  
EMPIRICAL POWERS FOR  $\alpha = .10$ ,  $p = 3$ ,  $n = 10$

	$\underline{W}_p$	$\underline{W}_{\min}$	$\underline{u}_{\min}, \underline{u}_{\max}$	$\underline{W}^*_{\min}$	$\underline{u}^*_{\min}, \underline{u}^*_{\max}$	$\underline{W}$	$\underline{u}$
LN	.64	.64	.18	.62	.14	.72	.13
U	.09	.11	.10	.14	.15	.15	.32
$t_{10}$	.16	.14	.11	.17	.14	.13	.13

### 9. CONCLUSIONS

The power of  $\underline{W}_{\min}$  was generally very close to that of  $\underline{W}_p$  over the range of bivariate distributions considered, except for the bivariate uniform distribution where the  $\underline{W}_p$  test appeared to be biased. The bivariate  $u$  test was much less powerful for skewed distributions, but superior for the bivariate uniform or normal-uniform. These results are not unexpected reflecting similar performances for  $W$  and  $u$  in the univariate case. In the trivariate case,  $\underline{W}_p$  and  $\underline{W}_{\min}$  were equivalent, and again superior to  $u$  for the lognormal, although for the uniform and  $t_{10}$  alternatives all tests had very low power for  $n = 10$ . The powers for the  $W^*$  and  $u^*$  tests based on polynomial mappings were very close to those of the tests based on hyperplane mappings, suggesting that the simpler polynomial mappings may be quite satisfactory.

I am grateful to L. J. Gleser and J. A. Hartigan for helpful comments, and to Dorothy Thayer for the programming. A program to obtain  $W_{\min}$  and  $u_{\min}, u_{\max}$  may be obtained from Mrs. Thayer, Division of Data Analysis and Research, Educational Testing Service, Princeton, N.J. 08540.

REFERENCES

- Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis. John Wiley, New York.
- Andrews, D. F. (1971). Plotting high dimensional data. 38th Session ISI, Contributed Papers. 22-26.
- D'Agostino, R. B. (1971). An omnibus test of normality for moderate and large size samples. Biometrika, 58, 341-8.
- David, H. A., Hartley, H. O., & Pearson, E. S. (1954). The distribution of the ratio, in a single normal sample, of range to standard deviation. Biometrika, 41, 482-93.
- Kowalski, C. J. (1970). The performance of some rough tests for bivariate normality before and after coordinate transformations to normality. Technometrics, 12, 517-44.
- Malkovich, J. F. & Afifi, A. A. (1971). On tests for multivariate normality. Unpublished.
- Nair, K. R. (1947). A note on the mean deviation from the median. Biometrika, 34, 360-2.
- Pearson, E. S. & Chandra Sekar, C. (1936). The efficiency of statistical tools and a criterion for the rejection of outlying observations. Biometrika, 28, 308-20.
- Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52, 591-611.
- Shapiro, S. S., Wilk, M. B., & Chen, H. J. (1968). A comparative study of various tests for normality. J. Am. Statist. Ass. 63, 1343-72.