The Annals of Statistics
1982, Vol. 10, No. 3, 943–954

DETECTION OF MULTIVARIATE NORMAL OUTLIERS

By Steven J. Schwager and Barry H. Margolin

Cornell University and N.I.E.H.S.

The general outlier problem for a multivariate normal random sample with mean slippage is defined and is shown to be invariant under a natural group of transformations. A family of maximal invariants is obtained, and the common distribution of its members is derived. The critical region for the locally best invariant test of the null hypothesis, that there are no outliers, versus the alternative hypothesis, that some outliers are present, is found. Under very general conditions, this test is equivalent to rejecting the null hypothesis whenever Mardia's multivariate sample kurtosis is sufficiently large.

1. Introduction. Anscombe and Tukey (1963, page 146) considered outliers to be "observations that have such large residuals, in comparison with most of the others, as to suggest that they ought to be treated specially." These aberrant observations can result from various underlying conditions, including model inadequacies and occurrences of gross observational errors. To propose and compare outlier procedures, one must know what information is sought from the analysis. As Kruskal (1960) and Gnanadesikan (1977, page 272) have noted, an observation may be an outlier for one purpose but not for another. Two possible aims were mentioned by David (1981, page 218): (a) to determine whether outliers are present in the data, and (b) to identify those observations that are aberrant. Clearly, if either or both of these are the objectives, the outliers themselves are the primary concern of the analysis. On the other hand, if fitting a model, estimating a set of parameters, or testing a hypothesis is the main interest, outliers are a complication, to be handled in an appropriate fashion. The aim there is: (c) to modify a statistical analysis, usually of a standard nature, by using information regarding the presence and identity of outliers. See, for example, Anscombe (1960). Methods suitable for one of these tasks may or may not be suitable for the others.

The focus here will be primarily on goal (a), outlier detection, for data that, if free of outliers, would be modeled as a random sample from a multivariate normal distribution. Any observation whose distribution departs from this model is an outlier. In the two models most widely used to represent the existence of outliers, all observations are normally distributed. Under the mean slippage model to be considered in this paper, all observations have a common covariance matrix Σ, but k of the means differ from the common mean μ of the rest, and possibly from each other. The variance slippage model is defined along similar lines, and will not be discussed in this paper.

The multivariate normal error structure has been adopted for several reasons, including mathematical tractability, and even more importantly, the fact that many of the standard multivariate methods are derived under the assumption of normality. This makes it crucial to check for outliers, as well as for other types of nonnormality, as their presence will strongly affect inferences made from normal-based procedures. For example, Layard (1974) showed that the normal theory likelihood ratio test for equality of covariance matrices is highly nonrobust against departures from normality, including contamination.

Most work on the outlier problem has been directed at the univariate case. This is easier to deal with than the multivariate case, as Gnanadesikan (1977, page 271) has pointed out:

Received April 1980; revised March 1982.

AMS 1970 *subject classifications.* Primary 62A05, 62H15; secondary 62C99, 62E15, 62H10.

Key words and phrases. Locally best invariant test, maximal invariant, mean slippage, multivariate kurtosis, multivariate normal distribution, outliers.

"The consequences of having defective reponses are intrinsically more complex in a multivariate sample than in the much-discussed univariate case. One reason is that a multivariate outlier can distort not only measures of location and scale but also those of orientation (i.e., correlation). A second reason is that it is much more difficult to characterize a multivariate outlier. A single univariate outlier may typically be thought of as 'the one that sticks out on the end,' but no such simple concept suffices in higher dimensions. A third reason is the variety of types of multivariate outliers that may arise; a vector response may be faulty because of a gross error in one of its components or because of systematic mild errors in all of its components."

Extensive surveys of the outlier literature are found in Barnett and Lewis (1978) and Hawkins (1980). Other general sources are David (1981) and Doornbos (1966). Gnanadesikan (1977) discussed multivariate outliers from a data analytic viewpoint. Various aspects of the multivariate outlier problem were treated by Siotani (1959), Karlin and Truax (1960), Ferguson (1961), Wilks (1963), Healy (1968), and Rohlf (1975).

The main result of this paper is that the locally best invariant test for outliers is based on Mardia's (1970) multivariate sample kurtosis $b_{2,p}$. Theorems 6.1 to 6.4, which state this precisely, establish the only optimality properties of $b_{2,p}$ known at present. The remainder of this paper is a derivation of this result and is organized as follows. The general outlier problem for a multivariate normal random sample with mean slippage is defined in Section 2, and is shown to be invariant with respect to a natural group of transformations. A family of maximal invariants with respect to this group is obtained and the common distribution of its members is derived in Section 3. The form of the critical region for the locally best invariant test of the null hypothesis, that there are no outliers, versus the alternative hypothesis, that some outliers are present, is found in Sections 4 and 5. Under very general conditions, it is shown in Section 6 that this test is equivalent to rejecting the null hypothesis whenever the multivariate sample kurtosis is sufficiently large.

2. The general outlier problem for a multivariate normal random sample. Consider a random sample from a multivariate normal distribution. The model for these data can be specified by the matrix equation $Y = e\mu + U$, where the $n \times p$ observation matrix Y has i.i.d. rows $Y_1, \cdots, Y_n$, e is an $n \times 1$ vector of 1's, μ is the unknown $1 \times p$ mean vector, and the rows of the $n \times p$ matrix U are i.i.d. $N(0, \Sigma)$ with covariance matrix Σ unknown. It will be assumed that $n \geq p + 1$ to insure that μ and Σ are estimable.

For any matrix $A = (a_{ij})$, define $\|A\| = (\sum_{ij} a_{ij}^2)^{1/2}$. To incorporate the possibility of outliers, the multivariate normal random sample model is embedded in a *multivariate mean model with mean slippage*:

$$Y = e\mu + \Delta^* A^* + U. \tag{2.1}$$

Here e, μ, and U are as above, and $n \geq p + 1$; furthermore, Δ^* is a nonnegative scalar, and A^* is an arbitrary $n \times p$ matrix such that: (C1) $\|A^*\| = 1$, unless $\Delta^* = 0$, in which case $A^* = 0$; and (C2) more than half of the rows of A^* are zero. In this model, the observation Y_i is an *outlier* if the ith row of A^* is nonzero. Equation (2.1) extends a univariate outlier model proposed by Ferguson (1961).

No outliers are present if and only if (iff) $\Delta^* = 0$. Condition (C2) requires that more than half of the observations are drawn from the $N(\mu, \Sigma)$ population; (C1), (C2), and the nonnegativity of Δ^* insure uniqueness of parametrization. The *general outlier problem* consists of model (2.1), hypotheses $H_0{:}\Delta^* = 0$ and $H_1{:}\Delta^* > 0$, action space $\mathscr{A} = \{D_0, D_1\}$, where D_i denotes the decision to act as if hypothesis H_i is true, $i = 0, 1$, state space $\Theta = \{(\Delta^*, A^*, \mu, \Sigma){:}\Sigma > 0, \Delta^* \geq 0, \text{(C1), (C2) hold}\}$, and loss function L with $L(\theta, D_i) = i$ if $\Delta^* = 0$, $L(\theta, D_i) = 1 - i$ if $\Delta^* > 0$.

When decision theory and invariance are discussed, notation and definitions will be consistent with Ferguson (1967).

It is clear from (C1) and (C2) that model (2.1) allows quite general configurations of outliers. The general outlier problem deals with a much broader class of outlier arrangements than the single outlier problem, in which it is specified that at most one outlier is present. The latter problem, which is commonly treated as having $n + 1$ alternative hypotheses and actions, is not dealt with in this paper, but see Schwager (1979).

Let $\mathscr{P}$ denote the group of all $n \times n$ permutation matrices, $G\ell(p)$ the group of $p \times p$ nonsingular matrices, $\mathbb{R}^p$ Euclidean p-space, and $\mathscr{Y}$ the space of $n \times p$ matrices. Consider the group $G = \mathscr{P} \times G\ell(p) \times \mathbb{R}^p$ with the group operation defined by

$$(\Gamma_2, C_2, c_2) \circ (\Gamma_1, C_1, c_1) = (\Gamma_2\Gamma_1, C_1C_2, c_1C_2 + c_2)$$

where $\Gamma_i \in \mathscr{P}$, $C_i \in G\ell(p)$, $c_i \in \mathbb{R}^p$ $(i = 1, 2)$. Then the general outlier problem is invariant under G, where the action is $g(Y) = \Gamma YC + ec$ for $g = (\Gamma, C, c) \in G$, and

$$\bar{g}(\theta) = \bar{g}(\Delta^*, A^*, \mu, \Sigma) = \begin{cases} (\Delta^* \| A^*C \|, \| A^*C \|^{-1}\Gamma A^*C, \mu C + c, C'\Sigma C) & \text{if } \Delta^* > 0, \\ (0, 0, \mu C + c, C'\Sigma C) & \text{if } \Delta^* = 0. \end{cases}$$

As the problem is invariant under G, only decision procedures invariant under G will be considered. Any such procedure must be a function of a maximal invariant with respect to G. A family of matrix-valued statistics, each member of which is maximally invariant under G, will be derived in the next section.

3. A family of maximal invariants with respect to G. The general outlier problem is invariant under permutation of the rows of Y, so if an ordering of the rows is specified, only functions of the ordered rows $Y_{(1)}, \cdots, Y_{(n)}$ need be considered. Invariance under addition of an arbitrary vector c to each row reduces consideration to functions of $Y_{(1)} - \bar{Y}, \cdots, Y_{(n)} - \bar{Y}$, where $\bar{Y}$ is the sample mean vector. Invariance under right multiplication of Y by any nonsingular matrix C suggests a matrix version of Ferguson's (1961) approach, which will now be developed; related work has been done by Butler (1981).

Under model (2.1), the matrix of residuals is $R = Y - e\bar{Y}$. Let $S = R'R$, and $M = I - (1/n)ee'$; M is $n \times n$ symmetric, idempotent, and positive semi-definite. Moreover, S is nonsingular and the n scalars $(Y_i - \bar{Y})S^{-1}(Y_i - \bar{Y})'$ are distinct with probability one. Reorder the rows of Y to make these scalars an increasing function of the index i, noting that neither S nor $\bar{Y}$ is affected by row permutations of Y. Let $\tilde{Y}$ denote the resulting matrix, and $Y_{(i)}$ the ith row of $\tilde{Y}$. Choose an arbitrary orthogonal $n \times n$ matrix P satisfying $P'MP = D$ where the $n \times n$ matrix $D = \text{diag}(1, 1, \cdots, 1, 0)$. Once a particular P is chosen, it is held fixed throughout the analysis. Let P^i denote the ith column of P, and P_i the ith row of P. Define an $(n - p - 1) \times n$ matrix Φ_1 and a $p \times n$ matrix Φ_2 by $\Phi_1 = (P^1 \cdots P^{n-p-1})'$ and $\Phi_2 = (P^{n-p} \cdots P^{n-1})'$. The $n \times p$ matrix $P'M\tilde{Y} = DP'\tilde{Y}$ may be partitioned as

$$\begin{bmatrix} \Phi_1\tilde{Y} \\ \Phi_2\tilde{Y} \\ 0 \end{bmatrix} \equiv \begin{bmatrix} V_1 \\ V_2 \\ 0 \end{bmatrix}. \tag{3.1}$$

Define the $(n - p - 1) \times p$ matrix-valued statistic

$$T(Y) = \begin{cases} \Phi_1\tilde{Y}(\Phi_2\tilde{Y})^{-1} = V_1V_2^{-1} & \text{if } V_2 \text{ is nonsingular,} \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

THEOREM 3.1. *$T(Y)$ is a maximal invariant with respect to G.*

Consideration of the distribution of $T[g^*(Y)]$, where g^* denotes $(I, \Sigma^{-1/2}, -\mu\Sigma^{-1/2})$, shows that, without loss of generality, one can set $\mu = 0$, $\Sigma = I$ in model (2.1), and simultaneously replace Δ^* by $\Delta = \Delta^* \| A^*\Sigma^{-1/2} \|$ and A^* by

$$A \equiv \begin{cases} A^*\Sigma^{-1/2} / \| A^*\Sigma^{-1/2} \| & \text{if } A^* \neq 0, \\ 0 & \text{if } A^* = 0. \end{cases}$$

Note that $\Delta = 0$ iff $\Delta^* = 0$, and that A and A^* have exactly the same nonzero rows.

To obtain the distribution of T under model (2.1), define the $p \times p$ matrix $Q = T'T + I$, and let a_i and $\bar{a}$ denote the ith row and the row mean of A, respectively. Summation over the set of all permutations σ of the first n positive integers will be denoted by $\sum_\sigma^*$, and summation with index i ranging from 1 to n by $\sum_i$.

Summing the density of $Y_{\sigma(1)}, \cdots, Y_{\sigma(n)}$ over all permutations σ gives the density of $Y_{(1)}, \cdots, Y_{(n)}$ as

$$f_{\tilde{Y}}(Y_1, \cdots, Y_n) = \textstyle\sum_\sigma^* f_\sigma(Y_1, \cdots, Y_n)$$
$$= (2\pi)^{-np/2}\exp[-\frac{1}{2}(\textstyle\sum_i Y_i Y_i' + \Delta^2 \sum_i a_i a_i')] \sum_\sigma^* \exp[\Delta \sum_i Y_i a_{\sigma(i)}']$$

for the region where $Y_1, \cdots, Y_n$ make the scalars $(Y_i - \bar{Y})S^{-1}(Y_i - \bar{Y})'$, $i = 1, \cdots, n$ an increasing sequence; the density is zero elsewhere.

Since $P'MP = D$, eigenvector methods establish that the last column of P is $P^n = n^{-1/2}e$. Define X by $n^{1/2}\bar{Y} = P^{n\prime}\tilde{Y}$, and V by $\begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$, so that

$$\begin{pmatrix} V \\ X \end{pmatrix} = P'\tilde{Y} \quad \text{and} \quad \tilde{Y} = P\begin{pmatrix} V \\ X \end{pmatrix}.$$

In changing variables from $\tilde{Y}$ to V, X, the Jacobian is $|\det P|^p = 1$; thus,

$$f_{V,X}(V, X) = (2\pi)^{-np/2}\exp\left[-\frac{1}{2}\left(\textstyle\sum_i P_i\begin{pmatrix} V \\ X \end{pmatrix}(V'\ X')P_i' + \Delta^2 \sum_i a_i a_i'\right)\right]$$
$$\cdot \textstyle\sum_\sigma^* \exp\left[\Delta \sum_i P_i \begin{pmatrix} V \\ X \end{pmatrix} a_{\sigma(i)}'\right].$$

Observe that

$$\textstyle\sum_i P_i\begin{pmatrix} V \\ X \end{pmatrix}(V'\ X')P_i' = \operatorname{tr}(V'V) + XX' \quad \text{and} \quad \sum_i P_i\begin{pmatrix} V \\ X \end{pmatrix}a_{\sigma(i)}' = \sum_i P_i\begin{pmatrix} V \\ 0 \end{pmatrix}a_{\sigma(i)}' + n^{1/2}X\bar{a}',$$

so that

$$f_{V,X}(V, X) = (2\pi)^{-np/2}\exp\left[-\frac{1}{2}\operatorname{tr}(V'V) - \frac{1}{2}XX' - \frac{1}{2}\Delta^2 \textstyle\sum_i a_i a_i'\right]$$
$$\cdot \textstyle\sum_\sigma^* \exp\left[\Delta \sum_i P_i\begin{pmatrix} V \\ 0 \end{pmatrix}a_{\sigma(i)}' + \Delta n^{1/2}X\bar{a}'\right].$$

To integrate out X, use the multivariate normal density, obtaining

$$f_V(V) = (2\pi)^{-(n-1)p/2}\exp\left[-\frac{1}{2}\operatorname{tr}(V'V) - \frac{1}{2}\Delta^2 \textstyle\sum_i (a_i - \bar{a})(a_i - \bar{a})'\right]$$
$$\cdot \textstyle\sum_\sigma^* \exp\left[\Delta \sum_i P_i\begin{pmatrix} V \\ 0 \end{pmatrix}a_{\sigma(i)}'\right].$$

With $T = V_1 V_2^{-1}$ as in (3.2), let $W = V_2$ and

$$J(n \times p) = \begin{bmatrix} T \\ I \\ 0 \end{bmatrix}, \quad \text{so} \quad \begin{bmatrix} V_1 \\ V_2 \\ 0 \end{bmatrix} = \begin{bmatrix} T \\ I \\ 0 \end{bmatrix} W = JW. \tag{3.3}$$

Changing variables from V to (T, W) and integrating out W gives the density of T. The Jacobian is $|\det \partial V/\partial(T, W)| = |\det W|^{n-p-1}$, and routine substitution proves the following result:

THEOREM 3.2. *In the multivariate mean model with mean slippage, the density of the G-maximal invariant T is*

$$f_T(T) = (2\pi)^{-(n-1)p/2}\exp[-\frac{1}{2}\Delta^2 \textstyle\sum_i (a_i - \bar{a})(a_i - \bar{a})'] \times g(\Delta) \tag{3.4}$$

for the region where the scalars $P_iJQ^{-1}J'P_i'$, $i = 1, \cdots, n$ *form an increasing sequence, and is zero elsewhere. Here*

$$g(\Delta) = \textstyle\sum_\sigma^* \displaystyle\int \exp[-\frac{1}{2}\operatorname{tr}(W'QW) + \Delta \textstyle\sum_i P_iJWa'_{\sigma(i)}]\,|\det W|^{n-p-1}\,dW, \tag{3.5}$$

and the $p \times p$ *matrix* W *varies over all of* p^2*-dimensional space.*

Observe that, conditionally on A, $f_T(T)$ depends on the single parameter Δ. Consider A as given and fixed. This allows one to write the density of T as $f_T(T \mid \Delta)$ and to examine tests of $\Delta = 0$ versus $\Delta > 0$, conditional on knowledge of A. A particular test obtained through this conditioning process will be shown not to depend on A. It is therefore an unconditional test of $\Delta = 0$ against $\Delta > 0$.

4. The form of the critical region for invariant tests. Any nonrandomized test of $H_0{:}\Delta = 0$ versus $H_1{:}\Delta > 0$ that is invariant under G must be a function of T. The power function of such a test may be written in terms of the parameter Δ and critical region ω as

$$\beta_\omega(\Delta) = \int_\omega f_T(T \mid \Delta)\,dT. \tag{4.1}$$

The local behavior of a test at $\Delta = 0$ is determined by the derivatives of $\beta_\omega(\Delta)$ at $\Delta = 0$. Let k denote the smallest positive integer such that $\beta_\omega^{(k)}(0)$ is not identically zero for all ω. The locally best test of $\Delta = 0$ against $\Delta > 0$ can be found by maximizing $\beta_\omega^{(k)}(0)$ over the class of α-level tests. Two distinct cases occur regarding locally best unbiased tests of $\Delta = 0$ against $\Delta \neq 0$. If k is even, the locally best test is also locally best unbiased whenever it is unbiased. If k is odd, the locally best unbiased test can be found by maximizing $\beta_\omega^{(k+1)}(0)$ subject to the conditions $\beta_\omega(0) = \alpha$, and unbiasedness.

For the general outlier problem, β_ω is an even function of Δ for any invariant test ω, i.e., $\beta_\omega(\Delta) = \beta_\omega(-\Delta)$ for all Δ, ω. To observe this, note that the transformation $H = -W$ in each integral of (3.5) yields $g(\Delta) = g(-\Delta)$; this, (3.4), and (4.1) complete the demonstration. The power curve of any invariant test is thus symmetric with respect to the β_ω-axis, and has first derivative zero at $\Delta = 0$. This is related to the problem's invariance under G.

For any nonnegative integer j, define

$$v_j(T) = \frac{\partial^j}{\partial\Delta^j} f_T(T \mid \Delta)\,|_{\Delta=0}, \qquad \mathscr{F}_j = \textstyle\sum_\sigma^* [\sum_{i=1}^n P_iJWa'_{\sigma(i)}]^j. \tag{4.2}$$

Derivatives of the power function $\beta_\omega(\Delta)$ at $\Delta = 0$ can be computed from (4.1). This is facilitated by the interchange of differentiation and integration, which can be justified by the Dominated Convergence Theorem, as proved in Schwager (1979). In the course of this proof, it is shown that derivatives of $g(\Delta)$ can be obtained by differentiating under the integral sign in (3.5). In other words, for all integers $j \geq 0$,

$$\frac{\partial^j}{\partial\Delta^j} \beta_\omega(\Delta)\,|_{\Delta=0} = \int_\omega v_j(T)\,dT, \tag{4.3}$$

$$g^{(j)}(0) = \int \mathscr{F}_j \operatorname{etr}(-\frac{1}{2} W'QW)\,|\det W|^{n-p-1}\,dW. \tag{4.4}$$

Furthermore, $g^{(j)}(0) = 0$ for all odd values of j, since the integrand in (4.4) is an odd function of W.

Recall that $Q = T'T + I$. Given T, there exist a $p \times p$ orthogonal matrix K and a $p \times p$ diagonal matrix E such that $Q = K'EK$. Define

$$Z(p \times p) = Q^{1/2}W = K'E^{1/2}KW;$$

$$\begin{aligned}(4.5)\qquad &\gamma_i(p\times 1)=(a_i-\bar{a})', \qquad i=1,\cdots,n;\ \text{ and}\\ &r_i(1\times p)=P_iJQ^{-1/2}=P_iJK'E^{-1/2}K, \qquad i=1,\cdots,n.\end{aligned}$$

The following results (Schwager, 1979) are employed in evaluating the derivatives of (4.3):

$$\begin{aligned}&\text{(i) } W'QW=Z'Z; \quad \text{(ii) } |\det \partial W/\partial Z| = (\det Q)^{-p/2};\\ (4.6)\qquad &\text{(iii) } \textstyle\sum_{i=1}^n r_i=0 \ \text{ and } \ \sum_{i=1}^n r_i'r_i=I; \quad \text{(iv) } \sum_{i=1}^n \gamma_i=0;\\ &\text{(v) for any permutation } \sigma, \textstyle\sum_{i=1}^n P_iJWa'_{\sigma(i)}=\sum_{i=1}^n r_iZ\gamma_{\sigma(i)}; \ \text{ and}\\ &\text{(vi) } \mathscr{F}_2=(n-2)!n\textstyle\sum_{i=1}^n \gamma_i'Z'Z\gamma_i=(n-2)!n\sum_{i=1}^n\gamma_i'W'QW\gamma_i.\end{aligned}$$

LEMMA 4.1. $v_1(T)=v_2(T)=v_3(T)=0$ *for all* T.

PROOF. Define $\mathscr{D}_j\equiv(\partial^j/\partial\Delta^j)\log f_T(T\mid\Delta)\mid_{\Delta=0}$. It is immediate from (4.2) that

$$\begin{aligned}&\mathscr{D}_1=v_1(T)/v_0(T), \qquad \mathscr{D}_2=v_2(T)/v_0(T)-[v_1(T)/v_0(T)]^2,\\ &\mathscr{D}_3=v_3(T)/v_0(T)-3v_1(T)v_2(T)/[v_0(T)]^2+2[v_1(T)/v_0(T)]^3.\end{aligned}$$

Thus, it suffices to establish that $\mathscr{D}_j=0$ for all T, for $j=1, 2, 3$. It follows from (3.4) and $g'(0)=g^{(3)}(0)=0$ that $\mathscr{D}_1=\mathscr{D}_3=0$ for all T,

$$(4.7)\qquad \mathscr{D}_2=-\textstyle\sum_{i=1}^n\gamma_i'\gamma_i+g''(0)/g(0) \quad\text{and}\quad \mathscr{D}_4=[g^{(4)}(0)/g(0)]-3[g''(0)/g(0)]^2.$$

Define the integrals over p^2-dimensional Euclidean space

$$\begin{aligned}\psi_0&=\int \operatorname{etr}(-\tfrac{1}{2}Z'Z)(\det Z'Z)^{(n-p-1)/2}\,dZ,\\ \psi_i&=\int \gamma_i'Z'Z\gamma_i\operatorname{etr}(-\tfrac{1}{2}Z'Z)(\det Z'Z)^{(n-p-1)/2}\,dZ, \qquad i=1,\cdots,n.\end{aligned}$$

Then (4.4), (4.5), and (4.6) show that

$$\begin{aligned}(4.8)\qquad &g(0)=n!(\det Q)^{-(n-1)/2}\psi_0,\\ &g''(0)=(n-2)!n(\det Q)^{-(n-1)/2}\textstyle\sum_{i=1}^n\psi_i.\end{aligned}$$

It is helpful to reexpress ψ_0 and ψ_i as integrals over the space S_p^+ of $p\times p$ positive definite symmetric matrices. Background material may be found in Eaton (1972, Chapters 6 and 8). Let Z have density $p(Z)/\psi_0$, and let $S=Z'Z$. It follows from Eaton (Proposition 8.105) that S has a Wishart$(I, p, n-1)$ distribution. This is also the distribution of $\Lambda'S\Lambda$ for any $p\times p$ orthogonal matrix Λ. Choose Λ so that $\Lambda\gamma_i/\|\gamma_i\|$ is the unit vector $(1, 0, 0, \cdots, 0)'$. The (1,1) element s_{11} of S has mean $n-1$, so

$$\begin{aligned}\psi_i&=\psi_0E[\gamma_i'Z'Z\gamma_i]=\psi_0E[\gamma_i'S\gamma_i]\\ &=\psi_0\gamma_i'\gamma_iE[(\gamma_i'\Lambda'/\|\gamma_i\|)S(\Lambda\gamma_i/\|\gamma_i\|)]\\ &=\psi_0\gamma_i'\gamma_iE[s_{11}]=\psi_0\gamma_i'\gamma_i(n-1), \qquad i=1,\cdots,n.\end{aligned}$$

Thus $g''(0)/g(0)=\sum_{i=1}^n\gamma_i'\gamma_i$, establishing that $\mathscr{D}_2=0$. □

THEOREM 4.1. *For* $j=1, 2, 3$, $(\partial^j/\partial\Delta^j)\beta_\omega(\Delta)\mid_{\Delta=0}=0$ *for any (invariant) region* ω; $(\partial^4/\partial\Delta^4)\beta_\omega(\Delta)\mid_{\Delta=0}$ *is maximized by a region of the form*

$$(4.9)\qquad \omega=\{T: g^{(4)}(0)/g(0)\ge k_0\},$$

where the constant k_0 *is determined by the size of the test.*

PROOF. The first assertion follows from (4.3) and Lemma 4.1. It is routine to show that $\mathscr{D}_4=v_4(T)/v_0(T)$. The Generalized Neyman-Pearson Lemma, Lemma 4.1, and (4.7) show that the region maximizing $(\partial^4/\partial\Delta^4)\beta_\omega(\Delta)\,|_{\Delta=0} = \int_\omega v_4(T)\, dT$ is

$$\omega = \{T: v_4(T) \geq k_0 v_0(T) + \cdots + k_3 v_3(T)\} = \{T: v_4(T) \geq k_0 v_0(T)\}$$
$$= \{T: [g^{(4)}(0)/g(0)] - 3[g''(0)/g(0)]^2 \geq k_0\}.$$

Conditional on the a_i's, the term $3[g''(0)/g(0)]^2$ is a constant, since $g''(0)/g(0) = \sum_{i=1}^n \gamma_i'\gamma_i$. It can therefore be absorbed into k_0. □

5. Evaluation of $g^{(4)}(0)$. A change of variables from W to Z in (4.4) yields

$$g^{(4)}(0) = (\det Q)^{-(n-1)/2} \int \mathcal{F}_4 \operatorname{etr}\left(-\frac{1}{2} Z'Z\right) |\det Z|^{n-p-1}\, dZ.$$

A useful expression for $\mathcal{F}_4$ follows.

THEOREM 5.1. $\mathcal{F}_4 = (n-4)![(n^3+n^2)\mathcal{S}_1 + (3n^2-9n+3)\mathcal{S}_2 - (3n^2-3n-6)\mathcal{S}_3 - (3n^2-3n)\mathcal{S}_4 + 6\mathcal{S}_5]$,

where

$$\mathcal{S}_1 = \textstyle\sum_{i,j=1}^n (r_i Z\gamma_j)^4, \qquad \mathcal{S}_2 = (\sum_{i=1}^n \gamma_i' Z'Z\gamma_i)^2, \qquad \mathcal{S}_3 = \sum_{i=1}^n (\gamma_i' Z'Z\gamma_i)^2,$$
$$\mathcal{S}_4 = \textstyle\sum_{i=1}^n [\sum_{j=1}^n (r_i Z\gamma_j)^2]^2, \quad \text{and} \quad \mathcal{S}_5 = \sum_{i,j=1, i\neq j}^n (\gamma_i' Z'Z\gamma_j)^2.$$

PROOF. Let $x_{\sigma,i} = r_i Z\gamma_{\sigma(i)}$. It follows from the multinomial theorem and result (4.6)(v) that

$$\mathcal{F}_4 = \textstyle\sum_\sigma^* [\sum_i x_{\sigma,i}]^4 = \sum_\sigma^* \{\sum_i x_{\sigma,i}^4 + 4\sum_{ij\neq} x_{\sigma,i}^3 x_{\sigma,j} + 3\sum_{ij\neq} x_{\sigma,i}^2 x_{\sigma,j}^2 + \cdots\}. \tag{5.1}$$

Each sum on the right can be expressed in terms of $\mathcal{S}_1$ to $\mathcal{S}_5$ by repeated use of results (4.6)(iii) and (iv). Hence

(i) $$\textstyle\sum_\sigma^* \sum_i x_{\sigma,i}^4 = (n-1)! \sum_{i,i'} (r_i Z\gamma_{i'})^4 = (n-1)!\mathcal{S}_1.$$

(ii) $$\textstyle\sum_\sigma^* \sum_{ij\neq} x_{\sigma,i}^3 x_{\sigma,j} = (n-2)! \sum_{ij\neq, i'j'\neq} (r_i Z\gamma_{i'})^3 (r_j Z\gamma_{j'})$$
$$= (n-2)! \textstyle\sum_{i,i'} (r_i Z\gamma_{i'})^3 (\sum_{j\neq i} r_j) Z (\sum_{j'\neq i'} \gamma_{j'}) = (n-2)!\mathcal{S}_1.$$

The derivation of the remaining three sums is technically involved but similar (Schwager, 1979). Substituting into (5.1) produces

$$\mathcal{F}_4 = (n-1)!\mathcal{S}_1 + 4\cdot(n-2)!\mathcal{S}_1 + 3\cdot(n-2)!(\mathcal{S}_1 + \mathcal{S}_2 - \mathcal{S}_3 - \mathcal{S}_4) +$$
$$6\cdot(n-3)!\cdot(4\mathcal{S}_1 + \mathcal{S}_2 - 2\mathcal{S}_3 - 2\mathcal{S}_4) + (n-4)!(36\mathcal{S}_1 + 3\mathcal{S}_2 - 12\mathcal{S}_3 - 18\mathcal{S}_4 + 6\mathcal{S}_5).$$

Regrouping terms gives the expression in the statement of the theorem. □

COROLLARY 5.1.

$$g^{(4)}(0) = (n-4)!(\det Q)^{-(n-1)/2} \int [(n^3+n^2)\mathcal{S}_1 + (3n^2-9n+3)\mathcal{S}_2$$
$$- (3n^2-3n-6)\mathcal{S}_3 - (3n^2-3n)\mathcal{S}_4 + 6\mathcal{S}_5] \operatorname{etr}\left(-\frac{1}{2} Z'Z\right) |\det Z|^{n-p-1}\, dZ,$$

where the region of integration is all of p^2-dimensional space.

Only two of the integrals $\int \mathcal{S}_i \operatorname{etr}(-½Z'Z)\,|\det Z|^{n-p-1}\, dZ$, $i = 1, \cdots, 5$, must be calculated, for conditional on A, the integrals with leading terms $\mathcal{S}_2$, $\mathcal{S}_3$, and $\mathcal{S}_5$ are constants. Let

$$k_1 = \int [(3n^2 - 9n + 3)\mathscr{S}_2 - (3n^2 - 3n - 6)\mathscr{S}_3 + 6\mathscr{S}_5]\operatorname{etr}\left(-\frac{1}{2}Z'Z\right)|\det Z|^{n-p-1}\,dZ.$$

The following lemmas are needed to evaluate the integrals with leading terms $\mathscr{S}_1$ and $\mathscr{S}_4$.

LEMMA 5.1. *Define the constant* Φ, *depending only on* n *and* p, *by*

$$\Phi = \int z_{11}^4 \operatorname{etr}\left(-\frac{1}{2}Z'Z\right)|\det Z|^{n-p-1}\,dZ,$$

where the $p \times p$ *matrix* Z *is integrated over* p^2*-dimensional space. Then for any* $1 \times p$ *row vector* r *and* $p \times 1$ *column vector* c,

$$\int (rZc)^4 \operatorname{etr}\left(-\frac{1}{2}Z'Z\right)|\det Z|^{n-p-1}\,dZ = \|r\|^4\|c\|^4\,\Phi.$$

PROOF. Define $\hat{r} = r/\|r\|$ and $\hat{c} = c/\|c\|$, and choose orthogonal $p \times p$ matrices R and C such that $\hat{r}$ is the first row of R, and $\hat{c}$ the first column of C. Define a $p \times p$ matrix variable $X = RZC$, so $x_{11} = \hat{r}Z\hat{c}$. Then $rZc = \|r\|\|c\|x_{11}$, and changing variables from Z to X completes the proof. □

COROLLARY 5.2.

$$\int \mathscr{S}_1 \operatorname{etr}\left(-\frac{1}{2}Z'Z\right)|\det Z|^{n-p-1}\,dZ = [\textstyle\sum_{i=1}^n \|r_i\|^4][\sum_{i=1}^n \|\gamma_i\|^4]\Phi.$$

LEMMA 5.2.

$$\int z_{11}^2 z_{12}^2 \operatorname{etr}\left(-\frac{1}{2}Z'Z\right)|\det Z|^{n-p-1}\,dZ = \frac{1}{3}\Phi.$$

PROOF. For any $p \times p$ orthogonal matrix P with first column P_1, a change of variables to $X = ZP$ shows that

$$\Phi = \int [(x_{11} \cdots x_{1p})P_1]^4 \operatorname{etr}\left(-\frac{1}{2}X'X\right)|\det X|^{n-p-1}\,dX. \tag{5.2}$$

Multiply both sides of the identity

$$[2^{-1/2}(z_{11} + z_{12})]^4 + [2^{-1/2}(z_{11} - z_{12})]^4 = \tfrac{1}{2}(z_{11}^4 + 6z_{11}^2 z_{12}^2 + z_{12}^4)$$

by $\operatorname{etr}(-\tfrac{1}{2}Z'Z)|\det Z|^{n-p-1}$ and integrate with respect to dZ, using (5.2) with P_1 equal to $(2^{-1/2}, \pm 2^{-1/2}, 0, \cdots, 0)'$ and $(0, 1, 0, \cdots, 0)'$. □

THEOREM 5.2.

$$\int \mathscr{S}_4 \operatorname{etr}\left(-\frac{1}{2}Z'Z\right)|\det Z|^{n-p-1}\,dZ = \frac{1}{3}\Phi[\textstyle\sum_{i=1}^n \|r_i\|^4][2\sum_{j,k=1}^n (\gamma_j'\gamma_k)^2 + (\sum_{j=1}^n \|\gamma_j\|^2)^2].$$

PROOF. For any i, by simultaneous diagonalization (see Press, 1972, page 37), there exists an orthogonal $p \times p$ matrix U such that $U(r_i'r_i)U' = \operatorname{diag}(\|r_i\|^2, 0, 0, \cdots, 0)$. Define a $p \times p$ matrix variable X by $X = UZ$, so that

$$\textstyle\sum_{j=1}^n \gamma_j'Z'r_i'r_iZ\gamma_j = \sum_{j=1}^n \gamma_j'X' \operatorname{diag}(\|r_i\|^2, 0, \cdots, 0)X\gamma_j = \|r_i\|^2[x_{11} \cdots x_{1p}]C[x_{11} \cdots x_{1p}]',$$

where the $p \times p$ matrix $C = \sum_{j=1}^n \gamma_j\gamma_j'$. Defining

$$\Pi = \int \{[x_{11} \cdots x_{1p}]C[x_{11} \cdots x_{1p}]'\}^2 \operatorname{etr}\left(-\frac{1}{2}X'X\right)|\det X|^{n-p-1}\,dX,$$

it follows that

$$\int [\sum_{j=1}^{n} \gamma_j' Z' r_i' r_i Z \gamma_j]^2 \operatorname{etr}\left(-\frac{1}{2} Z'Z\right) |\det Z|^{n-p-1}\, dZ = \|r_i\|^4\, \Pi.$$

Since Π is independent of the index i, the definition of $\mathcal{S}_4$ shows that

$$\int \mathcal{S}_4 \operatorname{etr}\left(-\frac{1}{2} Z'Z\right) |\det Z|^{n-p-1}\, dZ = [\sum_{i=1}^{n} \|r_i\|^4]\, \Pi.$$

Again by simultaneous diagonalization, there exists an orthogonal $p \times p$ matrix V such that $VCV' = \operatorname{diag}(\lambda_1, \cdots, \lambda_p)$ where $\lambda_1, \cdots, \lambda_p$ are the eigenvalues of C. Define a $p \times p$ matrix variable by $Y = XV'$. Then

$$\Pi = \int [\sum_{i=1}^{p} \lambda_i y_{1i}^2]^2 \operatorname{etr}\left(-\frac{1}{2} Y'Y\right) |\det Y|^{n-p-1}\, dY$$

$$= [\sum_{i=1}^{p} \lambda_i^2]\Phi + [\sum_{i,j=1, i\neq j}^{p} \lambda_i\lambda_j] \frac{1}{3}\Phi = \frac{1}{3}\Phi[2\sum_{i=1}^{p} \lambda_i^2 + (\sum_{i=1}^{p} \lambda_i)^2].$$

It suffices now to use the relations

$$\sum_{i=1}^{p} \lambda_i = \operatorname{tr} C = \sum_{j=1}^{n} \|\gamma_j\|^2, \qquad \sum_{i=1}^{p} \lambda_i^2 = \operatorname{tr}(C^2) = \sum_{j,k=1}^{n} (\gamma_j'\gamma_k)^2$$

to complete the proof. □

Theorem 5.3 summarizes the derivation of $g^{(4)}(0)$.

THEOREM 5.3. *The derivative $g^{(4)}(0)$ is given by*

$$g^{(4)}(0) = (n-4)!(\det Q)^{-(n-1)/2}[\Phi L \sum_{i=1}^{n} \|r_i\|^4 + k_1], \tag{5.3}$$

where the constant k_1 depends on A, n, and p, and

$$L = (n^3 + n^2) \sum_{i=1}^{n} \|\gamma_i\|^4 - (n^2 - n)[2\sum_{i,j=1}^{n} (\gamma_i'\gamma_j)^2 + (\sum_{i=1}^{n} \|\gamma_i\|^2)^2].$$

An associate editor has pointed out that an alternative method of deriving these results is based on the technique of expanding the density of any maximal invariant about $A = 0$. The expansion of the density can be obtained from Wijsman's Theorem (Wijsman, 1967).

6. Multivariate kurtosis and the locally best invariant test for the general outlier problem. Mardia (1970, 1974, 1975) has defined and treated the *multivariate sample kurtosis*

$$b_{2,p}(Y) \equiv b_{2,p} = n \sum_{i=1}^{n} [(Y_i - \bar{Y})S^{-1}(Y_i - \bar{Y})']^2.$$

THEOREM 6.1. *For the general outlier problem, the locally best invariant test of $H_0: \Delta = 0$ versus $H_1: \Delta > 0$, conditional on A, is: If $L > 0$, reject H_0 whenever $b_{2,p} \geq K$; if $L < 0$, reject H_0 whenever $b_{2,p} \leq K'$. The constants K and K' are determined by the size of the test, and L is the function of A given in Theorem* 5.3.

PROOF. From Theorem 4.3 and the discussion at the beginning of Section 4, the locally best invariant test is given by the critical region ω of (4.9). Substituting from (4.8) and (5.3) shows that ω is specified by

$$(n-4)![\Phi L \sum_i \|r_i\|^4 + k_1]/k_2 \geq k_0, \tag{6.1}$$

where k_0, k_1, and L depend on A, and k_2 and Φ are positive constants depending only on n and p. Absorbing constants into k_0 shows that (6.1) is equivalent to $L \sum_{i=1}^{n} \|r_i\|^4 \geq k_0'$, where L and k_0' are functions of A.

It follows from (3.1) and (3.3) that $P'M\tilde{Y} = JW$, and thus $\tilde{Y} - e\bar{Y} = M\tilde{Y} = PJW$. Since

$Y_i - \bar{Y} = P_iJW$ and $S = (\tilde{Y} - e\bar{Y})'(\tilde{Y} - e\bar{Y}) = W'QW$, it is immediate that $b_{2,p} = \sum_{i=1}^{n} \|r_i\|^4$. Thus, if L is positive, the region ω can be specified by $b_{2,p} = n \sum_i \|r_i\|^4 \geq nk_0'/L = K$; if L is negative, by $b_{2,p} = n \sum_i \|r_i\|^4 \leq nk_0'/L = K'$. □

The matrix A determines, through L, whether the locally best invariant test of Theorem 6.1 rejects H_0 when $b_{2,p}$ is too large or when it is too small. A related point is that if $L = 0$, (6.1) shows that the critical region of the locally best invariant test depends on the power function's derivatives of order greater than four. Both of these problems would be solved if it were known that $L > 0$ for all A of interest. Theorem 6.2 will show that L is positive whenever the fraction of nonzero rows of A is at most $(3 - \sqrt{3})/6 \simeq 21\%$. Theorem 6.3 will show that L is positive whenever $e'A = 0$, that is, the sum of the rows of A is 0, and at most one-third of the rows of A are nonzero.

THEOREM 6.2. *If $a_i = 0$ for $i = m + 1, \cdots, n$, and $m/n \leq (3 - \sqrt{3})/6 = .2113 \cdots$, then $L > 0$, and the test that rejects H_0 when $b_{2,p} \geq K$ is locally best invariant, uniformly in $(a_1, \cdots, a_m)$.*

PROOF. It must be shown that $L > 0$, or equivalently that

$$(n^3 + n^2) \textstyle\sum_i \|\gamma_i\|^4 > (n^2 - n)(2F + 1)(\sum_i \|\gamma_i\|^2)^2, \tag{6.2}$$

where

$$F = \textstyle\sum_{i,j} (\gamma_i'\gamma_j)^2/(\sum_i \|\gamma_i\|^2)^2.$$

Observing that F is nonnegative and summing the Cauchy-Schwarz inequality $(\gamma_i'\gamma_j)^2 \leq \|\gamma_i\|^2\|\gamma_j\|^2$ over all i and j, one sees that $0 \leq F \leq 1$. Consequently, it suffices for (6.2) to prove that

$$(n^3 + n^2) \textstyle\sum_i \|\gamma_i\|^4 > 3(n^2 - n)(\sum_i \|\gamma_i\|^2)^2. \tag{6.3}$$

In fact, this is also necessary for (6.2), as it is the special case of the latter obtained when all γ_i are scalar multiples of a common vector.

Two relations will prove useful. For $i > m$, $\gamma_i = -\bar{a}'$, so for any exponent k,

$$\textstyle\sum_{i=1}^{n} \|\gamma_i\|^k = \sum_{i=1}^{m} \|\gamma_i\|^k + (n - m)\|\bar{a}\|^k. \tag{6.4}$$

Also, taking the squared norm of each side in the identity $(n - m)\bar{a}' = \sum_{i=1}^{m} \gamma_i$ and repeatedly applying the inequality $\gamma_i'\gamma_j + \gamma_j'\gamma_i \leq \gamma_i'\gamma_i + \gamma_j'\gamma_j$ yields

$$(n - m)^2\|\bar{a}\|^2 \leq m \textstyle\sum_{i=1}^{m} \|\gamma_i\|^2.$$

Let x_i denote $\|\gamma_i\|^2$ for $i = 1, \cdots, m$, let y denote $\|\bar{a}\|^2$, and define $\bar{x} = m^{-1} \sum_{i=1}^{m} x_i$. It follows from (6.4) that (6.3) is equivalent to

$$G(x_1, \cdots, x_m, y) = \frac{[\sum_{i=1}^{m} x_i + (n - m)y]^2}{\sum_{i=1}^{m} x_i^2 + (n - m)y^2} < \frac{n^2 + n}{3(n - 1)}. \tag{6.5}$$

To examine the relationship between m and the maximum value of G on

$$\{(x_1, \cdots, x_m, y) : x_1, \cdots, x_m, y \geq 0;\ (n - m)^2 y \leq m^2\bar{x}\}, \tag{6.6}$$

begin by observing that G is increased by equalizing the x_i; that is,

$$G(x_1, \cdots, x_m, y) \leq G(\bar{x}, \cdots, \bar{x}, y) = [m\bar{x} + (n - m)y]^2/[m\bar{x}^2 + (n - m)y^2].$$

Because of its homogeneity, G may be treated as a function of the single variable $u = y/\bar{x}$. It is a routine exercise to find the maximum of $G(u) = [m + (n - m)u]^2/[m + (n - m)u^2]$ over the domain $u \in [0, m^2/(n - m)^2]$, for $G(u)$ is an increasing function, taking its maximum value $n^2m(n - m)/[m^3 + (n - m)^3]$ at $u = m^2/(n - m)^2$. It follows that a sufficient condition for (6.5) to hold over the region (6.6) is that

$$n^2m(n - m)/[m^3 + (n - m)^3] < (n^2 + n)/3(n - 1). \tag{6.7}$$

Let $s = m/n$; (6.7) holds whenever $s(1-s)/[s^3 + (1-s)^3] \leq 1/3$, or equivalently whenever $6s^2 - 6s + 1 \geq 0$. Solving this quadratic inequality, and restricting attention to the interval $(0, 1/2)$ as required by the model, gives

$$0 < s = m/n \leq (3 - \sqrt{3})/6 = .2113 \cdots. \square$$

THEOREM 6.3. *If $a_i = 0$ for $i = m+1, \cdots, n$, $\sum_{i=1}^{m} a_i = 0$, and $m/n \leq 1/3$, then $L > 0$, and the test that rejects H_0 when $b_{2,p} \geq K$ is locally best invariant, uniformly in $(a_1, \cdots, a_m)$.*

PROOF. With the notation of the last proof, $y = \| \bar{a} \|^2 = 0$, and letting X denote the $m \times 1$ vector $(x_1, \cdots, x_m)'$ and E the $m \times m$ matrix consisting entirely of ones, (6.5) becomes

$$G(X) = X'EX/X'X < (n^2 + n)/3(n-1). \tag{6.8}$$

The maximum of $X'EX/X'X$ is m, the largest eigenvalue of E, so (6.8) holds for all X when

$$m/n < (n+1)/3(n-1),$$

for which $m/n \leq 1/3$ is sufficient. $\square$

If the general outlier problem is assumed to have a fraction of outliers no greater than 21.13%, Theorem 6.2 gives a test for outliers that is locally best invariant for every A. If A is known to satisfy $e'A = 0$, this test remains locally best invariant when the fraction of outliers is as high as 33⅓%. Other restrictions could be placed on A, giving different bounds on the permissible fraction of outliers leading to the same result. However, this seems unnecessary in view of the large fraction of outliers for which the test based on $b_{2,p}$ is locally best for all A. It is interesting to note that this fraction does not depend on the dimension p of the observations.

Throughout this paper, the matrix A was assumed known. The multivariate kurtosis test was shown in Theorems 6.2 and 6.3 to be locally best invariant uniformly on all A's of certain types. A stronger result, which Ferguson has called strong local optimality, allows A to be unknown.

THEOREM 6.4. *Let ω be the critical region of Theorem 6.2, let ω' be the critical region of any other invariant test of the same size as ω, set $\Delta = 1$, and let k/n be less than $(3 - \sqrt{3})/6$, where k denotes the maximum number of nonzero rows of A. Assume that ω and ω' are distinct, meaning that the Lebesgue measure of their symmetric difference is positive. Then there exists a neighborhood of the origin in kp-dimensional space on which $\beta_\omega(a_1, \cdots, a_k) > \beta_{\omega'}(a_1, \cdots, a_k)$ except at the origin, where there is equality.*

The proof of this parallels the proof of a similar result in Ferguson (1961, Sec. 2.4), so details are omitted here. The discussion accompanying that result also applies to Theorem 6.4.

Relatively little guidance on how to search for multivariate normal outliers exists in print. Informal graphical procedures for discovering deviant observations have been proposed (Gnanadesikan, 1977). Another widely used technique is sequential one-at-a-time deletion of the "most outlying" observation (Hawkins, 1980). Both approaches have shortcomings.

The results of this section support the view that Mardia's multivariate sample kurtosis is sensitive to the presence of outliers as well as other distributional departures from normality. There is also empirical evidence that the kurtosis test behaves well in situations of practical interest when compared with other inferential outlier procedures. Obtaining evidence of the presence of an unspecified number of unidentified outliers does not answer the entire outlier question, but it constitutes an important beginning. Much like an F-test in an analysis of variance, it can serve as a license for subsequent outlier identification and adjustment.

The kurtosis test has the attractive feature that its local optimality holds whether the data contain one, two, or any number of outliers up to 21.13% of the observations. The misleading behavior of tests for a smaller number of outliers than are actually present, e.g., tests for two (or less) outliers in the presence of more than two outliers (Hawkins, 1980, page 57), increases the importance of performing an overall test that is sensitive to a broad range of outlier configurations.

Mardia (1970, 1974, 1975) has proposed that his multivariate sample kurtosis be employed as a test of nonnormality, with large values of kurtosis leading to rejection of the null hypothesis of normality. Considering the presence of a small number of outliers is a worthwhile alternative to searching through a family of transformations for a better fit to the data. The use of kurtosis as a preliminary screening device insures that whatever is done afterwards, e.g., sequential elimination or graphical inspection, cannot raise the overall type I error rate above that of the initial kurtosis test.

REFERENCES

ANSCOMBE, F. J. (1960). Rejection of outliers. *Technometrics* **2** 123–147.

ANSCOMBE, F. J. and TUKEY, J. W. (1963). The examination and analysis of residuals. *Technometrics* **5** 141–160.

BARNETT, V. and LEWIS, T. (1978). *Outliers in Statistical Data.* Wiley, New York.

BUTLER, R. W. (1981). The admissible Bayes character of subset selection techniques involved in variable selection, outlier detection, and slippage problems. *Ann. Statist.* **9** 960–973.

CONSTANTINE, A. G. (1963). Noncentral distribution problems in multivariate analysis. *Ann. Math. Statist.* **34** 1270–1285.

DAVID, H. A. (1981). *Order Statistics*, 2nd edition. Wiley, New York.

DOORNBOS, R. (1966). *Slippage Tests.* Mathematical Centre Tracts 15, Mathematisch Centrum, Amsterdam.

EATON, M. L. (1972). *Multivariate Statistical Analysis.* Inst. Math. Statist., Univ. Copenhagen, Copenhagen, Denmark.

FERGUSON, T. S. (1961). On the rejection of outliers. *Proc. Fourth Berkeley Symposium Math. Statist. Prob.* I, 253–287.

FERGUSON, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach.* Academic, New York.

GNANADESIKAN, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations.* Wiley, New York.

HAWKINS, D. M. (1980). *Identification of Outliers.* Chapman and Hall, New York.

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

JOHNSON, N. L. and KOTZ, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* Wiley, New York.

KARLIN, S. and TRUAX, D. R. (1960). Slippage problems. *Ann. Math. Statist.* **31** 296–324.

KRUSKAL, W. H. (1960). Some remarks on wild observations. *Technometrics* **2** 1–3.

LAYARD, M. W. J. (1974). A Monte Carlo comparison of tests for equality of covariance matrices. *Biometrika* **61** 461–465.

LOEVE, M. (1977). *Probability Theory I*, 4th Edition. Springer-Verlag, New York.

MARDIA, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57** 519–530.

MARDIA, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā B* **36** 115–128.

MARDIA, K. V. (1975). Assessment of multinormality and the robustness of Hotelling's T^2 test. *Appl. Statist.* **24** 163–171.

PRESS, S. J. (1972). *Applied Multivariate Analysis.* Holt, Rinehart and Winston, New York.

ROHLF, F. J. (1975). Generalization of the gap test for the detection of multivariate outliers. *Biometrics* **31** 93–101.

SCHWAGER, S. J. (1979). Detection of multivariate outliers. Ph.D. thesis, Yale Univ.

SIOTANI, M. (1959). The extreme value of the generalized distances of the individual points in the multivariate normal sample. *Ann. Inst. Statist. Math., Tokyo* **10** 183–208.

WIJSMAN, R. A. (1967). Cross-sections of orbits and their application to densities of maximal invariants. *Proc. Fifth Berkeley Symposium Math. Statist. Probability* **I**, 389–400.

WILKS, S. S. (1963). Multivariate statistical outliers. *Sankhyā A* **25** 406–427.

BIOMETRICS UNIT
CORNELL UNIVERSITY
337 WARREN HALL
ITHACA, NEW YORK 14853

NATIONAL INSTITUTE OF ENVIRONMENTAL HEALTH SCIENCES
BIOMETRY AND RISK ASSESSMENT PROGRAM
MAIL DROP A3-01
RESEARCH TRIANGLE PARK, NORTH CAROLINA 22709