BU-680-M

DETECTION OF MULTIVARIATE OUTLIERS

BY STEVEN J. SCHWAGER AND BARRY H. MARGOLIN
Cornell University and N.I.E.H.S.

Summary

The general outlier problem for a multivariate normal random sample with mean slippage is defined and 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 the multivariate sample kurtosis is sufficiently large.

DETECTION OF MULTIVARIATE OUTLIERS

BY STEVEN J. SCHWAGER AND BARRY H. MARGOLIN
Cornell University and N.I.E.H.S.

1. Introduction. Anscombe and Tukey (1963, p. 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. Anscombe (1960, p. 124) mentioned two sources of such gross errors: measurement error, the error in operating equipment and recording readings; and execution error, which encompasses any other "discrepancy between what we intend to do and what is actually done," such as measuring the wrong quantity or attribute.

No statistical procedure is completely immune to the debilitating effects of outliers, although some are far less sensitive than others. Close scrutiny of those observations having large residuals frequently leads to better understanding of the data and to a more nearly valid model.

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.

To sensibly propose and compare outlier procedures, one must know what information is sought from the analysis. As Kruskal (1960) and Gnanadesikan (1977, p. 272) have noted, an observation may be an outlier for one purpose but not for another. Two possible aims were mentioned by David (1970, p. 170): (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. Methods suitable for one of these tasks may or may not be suitable for the others.

As an example, consider a set of observations that are realizations of independent normal random variables with a common variance. It is believed that most of the observations have mean μ, but it is suspected that a few may have means that differ greatly from μ. Objective (a) could be achieved through an overall test of whether the observations as a group appear distributed as a normal sample or as a contaminated normal sample. Objective (b) requires a determination of exactly which observations show evidence of a mean shift. If the ultimate goal is to estimate μ, one way of meeting objective (c) would be to average the observations after downweighting or discarding those identified as outliers in (b). This type of estimation procedure has been investigated by Anscombe (1960).

To clarify the distinction between goals (a) and (b), outlier detection and identification, consider a univariate random sample from a mixture of normal distributions. The components N(0,1), N(2,1), and N(-2,1) of the mixture have mixing probabilities .90, .05, and .05, respectively. Identification of the outliers, the observations drawn from the component distributions with nonzero means, is an unreasonable goal. Nearly half of them, constituting about 5% of the sample, will lie in the interval (-2,2), while approximately 4.54% of the non-outliers, constituting about 4% of the sample, will lie outside this interval. Although this makes identifying the entire set of outliers a hopeless task, it is possible that an outlier detection test could strongly signal the presence of outliers somewhere in the data, without specifying their exact locations. Rejection of the null hypothesis of no outliers might thus lead to consideration of alternative models for the data.

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, p. 271) has pointed out:

> "The consequences of having defective responses 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."

Many outlier methods based on exclusively univariate techniques cannot be generalized to the multivariate setting. For instance, Dixon's r statistics, which are ratios of differences of order statistics, do not have an obvious multivariate version. These and other complexities, such as the greater difficulty of distributional computations, account for the scarcity of both theoretical results and practical tools for multivariate outlier problems.

An extensive survey of the outlier literature is found in Barnett and Lewis (1978). Other general sources are David (1970) 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 remainder of this paper 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 $Y_1, \cdots, Y_n$ 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 $u_1, \cdots, u_n$ 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 $||A|| = (\Sigma_{ij}a_{ij}^2)^{\frac{1}{2}}$. To incorporate the possibility of outliers, the multivariate normal random sample model is embedded in a <u>multivariate mean model with mean slippage</u>:

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

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 <u>outlier</u> if $a_i \neq 0$, where a_i denotes the i<u>th</u> row of A . Equation (2.1) extends a univariate outlier model proposed by Ferguson (1961).

No outliers are present and (2.1) reduces to a random sample model if and only if (iff) $\Delta = 0$. With this formulation, the study of outliers in a multivariate normal random sample becomes a one-parameter problem. Condition (C2) requires that more than half of the observations are drawn from the $N(\mu,\Sigma)$ population, in which no mean slippage is present. Conditions (C1) and (C2) and the restriction that Δ must be nonnegative insure uniqueness of parametrization (see Lemma 2.2).

Hypothesis testing for the presence or absence of outliers can be formulated in decision theoretic terms as the general outlier problem, which consists of (a) model (2.1); (b) hypotheses $H_0:\Delta=0$ and $H_1:\Delta>0$; (c) action space $\mathcal{A} = \{D_0, D_1\}$, where D_i denotes the decision to act as if hypothesis H_i is true, $i=0,1$; (d) state space $\Theta = \{\theta = (\Delta, A, \mu, \Sigma) : \Sigma > 0;\ \Delta \geq 0;$ (C1), (C2) hold$\}$; and (e) loss function L given by

$$L(\theta, D_i) = \begin{cases} i & \text{if} \quad \Delta = 0 \\ 1-i & \text{if} \quad \Delta > 0 \end{cases} .$$

Here and throughout this paper, when decision theory and invariance are discussed, notation and definitions will be consistent with Ferguson (1967).

A rule that chooses between D_0 and D_1 will be termed an outlier detection rule; it dictates a single overall decision on the presence or absence of outliers. Outlier identification rules, in contrast, attempt to determine the set of all outlying observations.

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.

The ensuing definitions aid in the treatment of invariance to follow. For any permutation $v = [v(1), \cdots, v(n)]$ of the first n positive integers, the permuted identity matrix corresponding to v, denoted by I^v, is an $n \times n$ matrix whose (j,k)th element is $\delta_{v(j),k}$, where δ is the

Kronecker delta. That is,

$$I^v(j,k) = \begin{cases} 1 & \text{if} \quad k = v(j) \\ 0 & \text{otherwise} \end{cases} \quad .$$

When postmultiplied by any vector, I^v has the effect of moving the $v(1)$th entry to the first position, ..., and the $v(n)$th entry to the nth position.

Let $\mathcal{Y}$ denote the space of $n \times p$ matrices. The transformation $g_{C,K,v}: \mathcal{Y} \to \mathcal{Y}$ is defined by

$$g_{C,K,v}(Y) = I^v(YC + eK) \quad ,$$

where C is a $p \times p$ nonsingular matrix, and K is an arbitrary $1 \times p$ vector. The set of all such transformations will be denoted by

$$G = \{g_{C,K,v}: C \text{ is } p \times p;\ |C| \neq 0;\ K \text{ is } 1 \times p\} \quad .$$

Throughout this paper, routine proofs, such as those of the next two lemmas, will be omitted.

LEMMA 2.1. *The set of transformations G is a group.*

LEMMA 2.2. *The parameter θ for the family of distributions in the general outlier problem is identifiable.*

These lemmas lead to the following.

THEOREM 2.1. *The general outlier problem is invariant under G. Furthermore, a decision rule d for this problem is invariant under G iff*

$$d[g_{C,K,v}(Y)] = d(Y) \text{ for all } Y \in \mathcal{Y},\ g_{C,K,v} \in G \quad .$$

PROOF. Choose $g = g_{C,K,v} \in G$ and $\theta = (\Delta, A, \mu, \Sigma) \in \Theta$. Let $Y \sim P_\theta$ indicate that Y is modeled by (2.1), subject to (C1) and (C2). Substitution shows that

$$g(Y) = I^V(YC + eK) = e(\mu C + K) + \Delta I^V AC + I^V UC \quad .$$

The rows of $I^V UC$ are i.i.d. $N(0, C'\Sigma C)$ random variables, $C'\Sigma C$ is positive definite, and AC is nonzero if A is. Also, more than half of the rows of AC are equal. Hence $g(Y) \sim P_{\bar{g}(\theta)}$, where

$$\bar{g}(\theta) = \begin{cases} (\Delta\|AC\|, \|AC\|^{-1} I^V AC, \mu C + K, C'\Sigma C) & \text{if} \quad \Delta > 0 \\ (0, 0, \mu C + K, C'\Sigma C) & \text{if} \quad \Delta = 0 \end{cases} \quad .$$

The uniqueness of $\bar{g}(\theta)$ follows from Lemma 2.2 (see Ferguson, 1967, p. 144), so the family $\{P_\theta, \theta \in \Theta\}$ is invariant under G; and for any $g \in G$ and $D_j \in \mathcal{A}$,

$$L(\theta, D_j) = L[\bar{g}(\theta), D_j] \qquad \text{for all } \theta \in \Theta \quad .$$

Setting $\tilde{g}(D_j) = D_j$ shows the invariance of the loss function under G . The second portion of the theorem now follows from the definition of an invariant decision rule (Ferguson, 1967, p. 148) and the nature of $\tilde{g}$. QED

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 K 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 (1977).

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 i<u>th</u> 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 i<u>th</u> column of P, and P_i the i<u>th</u> row of P . Define an $(n\text{-}p\text{-}1) \times n$ 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

$$(3.1)\qquad \begin{bmatrix} \Phi_1\tilde{Y} \\ \Phi_2\tilde{Y} \\ 0 \end{bmatrix} \equiv \begin{bmatrix} V_1 \\ V_2 \\ 0 \end{bmatrix} \quad .$$

Restrict attention to the subset $\{Y \in \mathcal{Y}: \Phi_2\tilde{Y}$ is nonsingular$\}$ of the sample space. Define the $(n-p-1) \times p$ matrix-valued statistic

$$(3.2)\qquad T(Y) = \Phi_1\tilde{Y}(\Phi_2\tilde{Y})^{-1} = V_1V_2^{-1} \quad .$$

REMARK. The subset of the sample space excluded from consideration in defining T is a set of measure zero in the situations to be investigated, and will therefore have no effect on the analysis. In principle, the maximal invariant could be extended to this set by considering which $p \times p$ submatrices of (3.1) are nonsingular, if any, and using statistics similar to T(Y) with these matrices replacing $\Phi_2\tilde{Y}$. However, as this is not necessary for the problems to be addressed here, it will not be discussed further.

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

PROOF. The proof appears in Schwager (1979).

The remainder of this section is devoted to obtaining the distribution of T under model (2.1). Let g^* denote $g_{C,K,v} \in G$ with $C = \Sigma^{-\frac{1}{2}}$, $K = -\mu\Sigma^{-\frac{1}{2}}$, and v the identity permutation. Then $g^*(Y) = Y\Sigma^{-\frac{1}{2}} - e\mu\Sigma^{-\frac{1}{2}}$, so $E[g^*(Y)] = \Delta A\Sigma^{-\frac{1}{2}}$, and the rows of $g^*(Y)$ are independent, each with $p \times p$ covariance matrix I . For the model (2.1), define

$$\Delta^* = \Delta\|A\Sigma^{-\frac{1}{2}}\|; \quad U^* = U\Sigma^{-\frac{1}{2}}; \quad \text{and } A^* = \begin{cases} A\Sigma^{-\frac{1}{2}}/\|A\Sigma^{-\frac{1}{2}}\| & \text{if } A \neq 0 \ , \\ 0 \quad \text{if } A = 0 & . \end{cases}$$

Since $\Delta^* = 0$ iff $\Delta = 0$, H_0 of the general outlier problem can be expressed as $\Delta^* = 0$, and H_1 as $\Delta^* > 0$. Simple algebra shows that

$$g^*(Y) = \Delta^* A^* + U^* \quad \text{and} \quad \|A^*\|^2 = 1 \text{ if } A^* \neq 0 \quad .$$

Thus, finding the distribution of $T[g^*(Y)]$ is equivalent to finding the distribution of $T(Y)$, and so, without loss of generality, one can take $\mu = 0$, $\Sigma = I$ in model (2.1).

Define the $p \times p$ matrix $Q = T'T + I$. Let the mean of the rows of A be denoted by $\bar{a}$. Summation over the set of all permutations σ of the first n positive integers will be denoted by Σ_σ^*, and summation with index i ranging from 1 to n by Σ_i.

For any such permutation σ, standard multivariate normal theory shows that the density of $Y_{\sigma(1)}, \cdots, Y_{\sigma(n)}$ is

$$f_\sigma(Y_1, \cdots, Y_n) = (2\pi)^{-\frac{1}{2}np} \exp\left[-\tfrac{1}{2}\left(\Sigma_i Y_i Y_i' - 2\Delta\Sigma_i Y_i a'_{\sigma(i)} + \Delta^2 \Sigma_i a_{\sigma(i)} a'_{\sigma(i)}\right)\right]$$

$$\text{for all } Y_1, \cdots, Y_n \quad .$$

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

(3.3)

$$\begin{aligned} f_{\tilde{Y}}(Y_1, \cdots, Y_n) &= \Sigma_\sigma^* f_\sigma(Y_1, \cdots, Y_n) \\ &= (2\pi)^{-\frac{1}{2}np} \exp\left[-\tfrac{1}{2}\left(\Sigma_i Y_i Y_i' + \Delta^2 \Sigma_i a_i a_i'\right)\right] \Sigma_\sigma^* \exp\left[\Delta \Sigma_i Y_i a'_{\sigma(i)}\right] \end{aligned}$$

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 $P^n = n^{-\frac{1}{2}}e$. Define X by $n^{\frac{1}{2}}\bar{Y} = P^{n\prime}\tilde{Y}$, and V by $\binom{V_1}{V_2}$, so

$$\binom{V}{X} = P'\tilde{Y} \quad \text{and} \quad \tilde{Y} = P\binom{V}{X} \quad .$$

In changing variables from $\tilde{Y}$ to V, X in (3.3), the Jacobian is $|\det P|^p = 1$; thus,

$$f_{V,X}(V,X) = (2\pi)^{-\frac{1}{2}np}\exp\left[-\tfrac{1}{2}\left(\Sigma_i P_i\binom{V}{X}(V' \;\; X')P_i' + \Delta^2\Sigma_i a_i a_i'\right)\right]$$
$$\cdot\Sigma_\sigma^* \exp\left[\Delta\Sigma_i P_i\binom{V}{X}a'_{\sigma(i)}\right] \quad .$$

Observe that

$$\Sigma_i P_i\binom{V}{X}(V' \;\; X')P_i' = \operatorname{tr}(V'V) + XX' \quad ,$$

and

$$\Sigma_i P_i\binom{V}{X}a'_{\sigma(i)} = \Sigma_i P_i\binom{V}{0}a'_{\sigma(i)} + n^{\frac{1}{2}}X\bar{a}' \quad ,$$

so that

$$f_{V,X}(V,X) = (2\pi)^{-\frac{1}{2}np}\exp\left[-\tfrac{1}{2}\operatorname{tr}(V'V) - \tfrac{1}{2}XX' - \tfrac{1}{2}\Delta^2\Sigma_i a_i a_i'\right]$$
$$\cdot\Sigma_\sigma^* \exp\left[\Delta\Sigma_i P_i\binom{V}{0}a'_{\sigma(i)} + \Delta n^{\frac{1}{2}}X\bar{a}'\right] \quad .$$

To integrate out X, rearrange terms and complete the square, noting that

$$\int \exp\left[-\tfrac{1}{2}XX' + \Delta n^{\frac{1}{2}}X\bar{a}' - \tfrac{1}{2}\Delta^2 n\bar{a}\bar{a}'\right] dX = (2\pi)^{\frac{1}{2}p} \quad ,$$

as the integrand is a multivariate normal density function, up to a constant. Therefore,

$$f_V(V) = (2\pi)^{-\frac{1}{2}(n-1)p}\exp\left[-\tfrac{1}{2}\operatorname{tr}(V'V) - \tfrac{1}{2}\Delta^2\Sigma_i(a_i - \bar{a})(a_i - \bar{a})'\right]$$
$$\cdot\Sigma_\sigma^* \exp\left[\Delta\Sigma_i P_i\binom{V}{0}a'_{\sigma(i)}\right] \quad .$$

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

$$(3.4) \qquad J(n\times p) = \begin{bmatrix} T \\ I \\ 0 \end{bmatrix}, \quad \text{so} \quad \begin{bmatrix} V_1 \\ V_2 \\ 0 \end{bmatrix} = \begin{bmatrix} T \\ I \\ 0 \end{bmatrix} W = JW \quad .$$

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

$$(3.5) \qquad f_T(T) = (2\pi)^{-\frac{1}{2}(n-1)p} \exp\left[-\tfrac{1}{2}\Delta^2 \Sigma_i (a_i - \bar{a})(a_i - \bar{a})'\right] \times g(\Delta)$$

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

$$(3.6) \qquad g(\Delta) = \Sigma_\sigma^* \int \exp\left[-\tfrac{1}{2}\mathrm{tr}(W'QW) + \Delta\Sigma_i P_i JWa_{\sigma(i)}'\right] |\det W|^{n-p-1} dW \quad ,$$

and the p x 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 $f_T(t|\Delta)$ 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. A nonrandomized test of $H_0:\Delta=0$ versus $H_1:\Delta>0$ for (2.1) is specified by a critical region ω. Any test invariant under G must be a function of T, so the power function of such a test may be written in terms of the underlying parameter Δ as

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

The local behavior of a test at $\Delta=0$ is determined by the derivatives of $\beta_\omega(\Delta)$ at $\Delta=0$. When the power function of any test allows the necessary derivatives, the locally best test of $\Delta=0$ against $\Delta>0$ is usually obtained as the α-level test that maximizes $\beta_\omega'(0)$, and the locally best unbiased test of $\Delta=0$ against $\Delta\neq 0$ as the unbiased α-level test that maximizes $\beta_\omega''(0)$. In the latter case, unbiasedness implies that $\beta_\omega'(0) = 0$.

If the first few derivatives of $\beta_\omega(\Delta)$ are zero at $\Delta=0$ for all tests, this approach can be extended to the first nonzero derivatives by a Taylor series argument. Let k denote the smallest positive integer such that $\beta_\omega^{(k)}(0)$ is not identically zero for all ω. The locally best test 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. 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 (3.6) yields $g(\Delta) = g(-\Delta)$; this, (3.5), 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.

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. The following consequence of the Dominated Convergence Theorem will be employed to justify this interchange.

THEOREM 4.1. (Loève, 1977, p. 127). If, for all Δ in a finite interval, $\partial F(x,\Delta)/\partial\Delta$ exists and $|\partial F(x,\Delta)/\partial\Delta| \le L(x)$ for some integrable function L, then for any Δ in the interval,

$$\frac{\partial}{\partial\Delta}\int F(x,\Delta)dx = \int \frac{\partial F(x,\Delta)}{\partial\Delta}\,dx \quad .$$

Let $N(\sigma)$, or simply N wherever possible, denote the $p \times p$ matrix

$$N(\sigma) = N = (a'_{\sigma(1)} \cdots a'_{\sigma(n)})PJ \quad .$$

Given T, and thus Q, there exist a $p \times p$ orthogonal matrix Γ and a $p \times p$ diagonal matrix E such that $Q = \Gamma' E \Gamma$. Define

$$\begin{aligned} Z(p\times p) &= Q^{\frac{1}{2}}W = \Gamma' E^{\frac{1}{2}}\Gamma W \, ; \\ Y_i(p\times 1) &= (a_i - \bar{a})' \qquad \text{for } i=1,\cdots,n \text{ ; and} \\ r_i(1\times p) &= P_i J Q^{-\frac{1}{2}} = P_i J\Gamma' E^{-\frac{1}{2}}\Gamma \qquad \text{for } i=1,\cdots,n \quad . \end{aligned} \tag{4.2}$$

The following lemma will be helpful in the application of Theorem 4.1.

LEMMA 4.1. For any permutation σ, no element of the $p \times p$ matrix $NQ^{-\frac{1}{2}}$ has absolute value greater than $D = n(p\bar{a}\bar{a}')^{\frac{1}{2}}$. This expression is independent of T.

PROOF. Let $\hat{A}$ denote the $p \times n$ matrix $[a'_{\sigma(1)} \cdots a'_{\sigma(n)}]$. Then

$$\mathrm{tr}(NQ^{-1}N') = \mathrm{tr}\left[\hat{A}'\hat{A}\begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}(r'_1 \cdots r'_n)\right]$$

$$= \Sigma_{i,j=1}^{n}(\hat{A}'\hat{A})_{ij} r_j r'_i \le \left(\max_{i,j} r_j r'_i\right)\Sigma_{i,j}(\hat{A}'\hat{A})_{ij} \quad .$$

The maximum value of $r_j r'_i$ occurs when $i = j$, and cannot exceed $\Sigma_{i=1}^{n} r_i r'_i$, which equals p since $\Sigma_{i=1}^{n} r'_i r_i = I$. Furthermore,

$$\Sigma_{i,j=1}^{n}(\hat{A}'\hat{A})_{ij} = e'\hat{A}'\hat{A}e = n^2\bar{a}\bar{a}' \quad ,$$

so $\mathrm{tr}(NQ^{-1}N') \le D^2$.

Noting that the sum of the squares of the elements of $NQ^{-\frac{1}{2}}$ is $\mathrm{tr}(NQ^{-1}N')$ and that D^2 is independent of T completes the proof. QED

THEOREM 4.2. Let ω be any invariant test for the general outlier problem. Then

$$(4.3) \qquad \frac{\partial^j}{\partial\Delta^j}\beta_\omega(\Delta)\Big|_{\Delta=0} = \frac{\partial^j}{\partial\Delta^j}\int_\omega f_T(T|\Delta)dT\Big|_{\Delta=0} = \int_\omega \frac{\partial^j}{\partial\Delta^j} f_T(T|\Delta)\Big|_{\Delta=0}dT \quad .$$

PROOF. Theorem 4.1 will be applied twice. The first application shows that derivatives of $g(\Delta)$ can be obtained from (3.6) by differentiating under the integral sign. For any permutation σ, the integrand of (3.6) can be written as

$$(4.4) \qquad F(W,\Delta) = \mathrm{etr}(-\tfrac{1}{2}W'QW + \Delta NW)|\det W|^{n-p-1} \quad .$$

Choose an arbitrary but fixed positive number B . It will now be shown that there exists an integrable function $L(W)$ such that, for all Δ in the interval $(-B,B)$, $|\partial F(W,\Delta)/\partial\Delta| \le L(W)$.

Let w_{ij} and n_{ij} denote the ijth entries of W and N . Then

$$|\mathrm{tr}\ NW| = |\Sigma_{i,j=1}^{p} n_{ji} w_{ij}| \le \Sigma_{i,j=1}^{p} |n_{ji}||w_{ij}| \quad ,$$

and

$$|\det W| \le \Sigma_{\nu}^{*} |w_{1,\nu(1)} \cdots w_{p,\nu(p)}| \quad ,$$

where ν ranges over all permutations of $(1, \cdots, p)$. It follows that $|\mathrm{tr}\ NW||\det W|^{n-p-1}$ is dominated by a finite linear combination $G(W)$ of absolute values of products of elements of W, so

$$(4.5) \qquad |\partial F(W,\Delta)/\partial\Delta| = \mathrm{etr}(-\tfrac{1}{2}W'QW + \Delta NW)|\mathrm{tr}\ NW||\det W|^{n-p-1}$$
$$\le \mathrm{etr}(-\tfrac{1}{2}W'QW + \Delta NW)G(W) \quad .$$

For any Δ in $(-B,B)$ and matrix W, $\Delta \mathrm{tr}(NW) \le |B\mathrm{tr}(NW)|$, so

$$\mathrm{etr}(\Delta NW) \le \mathrm{etr}(BNW) + \mathrm{etr}(-BNW) \quad .$$

Consequently,

$$(4.6) \qquad |\partial F(W,\Delta)/\partial\Delta| \le \mathrm{etr}(-\tfrac{1}{2}W'QW + BNW)G(W) + \mathrm{etr}(-\tfrac{1}{2}W'QW - BNW)G(W) \quad .$$

To establish that the right-hand side of (4.6) is integrable, note that it is the sum of two functions, both of which can be handled in the same fashion. The first of these can be rewritten as

$$(4.7) \qquad \mathrm{etr}[-\tfrac{1}{2}(W - BQ^{-1}N')'Q(W - BQ^{-1}N')]\mathrm{etr}(\tfrac{1}{2}B^2NQ^{-1}N')G(W) \quad .$$

The first term is, up to a constant involving only Q, the density function of p independent random variables, each of which is p-dimensional multivariate normal, and the second term is a constant. When (4.7) is integrated with respect to dW over p^2-dimensional space, each term of $G(W)$ results in an absolute moment of some product of elements of W . These moments are noncentral, as the distribution given by the first term is centered at $BQ^{-1}N'$. Each such moment is finite, so the integral of

(4.7) is too. This demonstrates the integrability of the right-hand side of (4.6), which can therefore serve as the function $L(W)$ in applying Theorem 4.1 to $F(W,\Delta)$ given by (4.4).

With this function, the equality of Theorem 4.1 holds for $\Delta = 0$, and in fact for any finite value of Δ, as B was arbitrary. An easy extension of this method shows that derivatives of any order can be passed under the integral sign; only the power of $|\text{tr } NW|$ in (4.5) requires change. Summing over all permutations σ demonstrates that derivatives of $g(\Delta)$ can be found by differentiating under the n! integral signs of (3.6).

The second application of Theorem 4.1 will establish (4.3). An integrable function of T that dominates $|\partial f_T(T|\Delta)/\partial\Delta|$ on an interval about the origin is needed. Let $c_0 = (2\pi)^{-\frac{1}{2}(n-1)p}$ and $c_1 = \Sigma_1^n(a_i - \bar{a})(a_i - \bar{a})' = \Sigma_1^n\gamma_i'\gamma_i$, so that

$$f_T(T|\Delta) = c_0 \exp(-\tfrac{1}{2}c_1\Delta^2)g(\Delta) \quad .$$

Then

$$\begin{aligned} |\partial f_T(T|\Delta)/\partial\Delta| &= c_0 \exp(-\tfrac{1}{2}c_1\Delta^2)|g'(\Delta) - c_1\Delta g(\Delta)| \\ &\le c_0[|g'(\Delta)| + Bc_1 g(\Delta)] \quad , \end{aligned}$$

where

$$g(\Delta) = \Sigma_\sigma^* \int \text{etr}[-\tfrac{1}{2}W'QW + \Delta N(\sigma)W]|\det W|^{n-p-1}dW \quad , \tag{4.8}$$

and

$$|g'(\Delta)| \le \Sigma_\sigma^* \int \text{etr}[-\tfrac{1}{2}W'QW + \Delta N(\sigma)W]|\text{tr}[N(\sigma)W]||\det W|^{n-p-1}dW \quad . \tag{4.9}$$

Performing a transformation of variables to Z of (4.2) changes each integral in (4.8) to

$$(\det Q)^{-\frac{1}{2}(n-1)}\int \exp[-\tfrac{1}{2}\text{tr}(Z'Z) + \Delta\text{tr}(NQ^{-\frac{1}{2}}Z)]|\det Z|^{n-p-1}dZ \quad . \tag{4.10}$$

Lemma 4.1 implies that for every matrix Z and $\Delta \epsilon (-B, B)$,

$$\Delta \mathrm{tr}(NQ^{-\frac{1}{2}}Z) = \Delta \Sigma_{i,j=1}^{p} (NQ^{-\frac{1}{2}})_{ji} Z_{ij} \leq BD\Sigma_{i,j=1}^{p} |Z_{ij}| \quad .$$

Let $\mathrm{sgn}_{\ell}(i,j)$, $i,j = 1, \cdots, p$, denote one of the 2^{p^2} possible arrangements of p^2 plus and minus signs, where ℓ is an index running from 1 to 2^{p^2} . For any Z, there exists an ℓ such that

$$\Sigma_{i,j=1}^{p} |Z_{ij}| = \Sigma_{i,j=1}^{p} \mathrm{sgn}_{\ell}(i,j) Z_{ij} \quad ,$$

so

$$\mathrm{etr}(\Delta NQ^{-\frac{1}{2}}Z) \leq \Sigma_{\ell} \exp\left[BD\Sigma_{i,j=1}^{p} \mathrm{sgn}_{\ell}(i,j) Z_{ij}\right] \quad .$$

The expression of (4.10) is then bounded by

$$(4.11) \qquad (\det Q)^{-\frac{1}{2}(n-1)} \Sigma_{\ell} \int \exp\left[-\tfrac{1}{2}\mathrm{tr}(Z'Z) + BD\Sigma_{i,j} \mathrm{sgn}_{\ell}(i,j) Z_{ij}\right] \cdot |\det Z|^{n-p-1} dZ \quad .$$

Each of these integrals is a finite constant relative to T, depending only on BD, and can be treated by the same method as the integral of (4.7).

Summing (4.10) over all permutations σ shows that

$$g(\Delta) \leq K_1 (\det Q)^{-\frac{1}{2}(n-1)} \quad ,$$

where K_1 is a constant independent of T . The integrals of (4.9) can be treated similarly, as each is bounded by an expression like (4.11), but with an extra term $D\,\Sigma_{i,j} |Z_{ij}|$, which bounds $\mathrm{tr}(NQ^{-\frac{1}{2}}Z)$, in the integrand. Thus

$$|g'(\Delta)| \leq K_2 (\det Q)^{-\frac{1}{2}(n-1)} \quad ,$$

where the constant K_2 does not depend on T, and

$$|\partial f_T(T|\Delta)/\partial\Delta| \le c_0(K_2 + Bc_1K_1)(\det Q)^{-\frac{1}{2}(n-1)} \quad .$$

Here integrability of the right-hand side follows immediately from the fact that $[\det(I+T'T)]^{-\frac{1}{2}(n-1)}$ is, up to a constant, the joint density function of the matrix T distribution (Press, 1972, p. 129).

Application of Theorem 4.1 demonstrates that the interchange of operations in (4.3) is valid for any finite Δ, as B was arbitrary, and for any measurable set ω . Again, the extension to higher derivatives requires only a modification of the constant $K_2 + Bc_1K_1$. QED

For any nonnegative integer j, define

$$\text{(4.12)} \qquad \begin{aligned} v_j(T) &= \frac{\partial^j}{\partial\Delta^j} f_T(T|\Delta)\Big|_{\Delta=0} \ ; \\ \mathfrak{J}_j &= \Sigma_\sigma^* \left[\Sigma_{i=1}^n P_i J W a'_{\sigma(i)}\right]^j \quad . \end{aligned}$$

COROLLARY 4.1. The size of any invariant region ω is $\beta_\omega(\Delta)\big|_{\Delta=0} = \int_\omega v_0(T)dT$, and

$$\text{(4.13)} \qquad \frac{\partial^j}{\partial\Delta^j}\beta_\omega(\Delta)\Big|_{\Delta=0} = \int_\omega v_j(T)dT, \qquad j=1,2,\cdots \quad .$$

The following lemmas will be used to evaluate the derivatives (4.13) with $1 \le j \le 4$.

LEMMA 4.2. For all integers $j \ge 0$,

$$\text{(4.14)} \qquad g^{(j)}(\Delta)\Big|_{\Delta=0} = \int \mathfrak{J}_j \operatorname{etr}(-\tfrac{1}{2}W'QW)|\det W|^{n-p-1}dW \quad .$$

PROOF. Take derivatives of $g(\Delta)$ by differentiating under the integral sign in (3.6), as discussed in the proof of Theorem 4.2. QED

LEMMA 4.3. $g^{(j)}(\Delta)\big|_{\Delta=0} = 0$ for all odd values of j .

PROOF. The integrand in (4.14) is an odd function of W . QED

LEMMA 4.4. For the variables defined in (4.2),

(i) $W'QW = Z'Z$;

(ii) $|\det \partial W/\partial Z| = (\det Q)^{-\frac{1}{2}p}$;

(iii) $\Sigma_{i=1}^n r_i = 0$ and $\Sigma_{i=1}^n r_i' r_i = I$;

(iv) $\Sigma_{i=1}^n \gamma_i = 0$;

(v) for any permutation σ, $\Sigma_{i=1}^n P_i JWa_{\sigma(i)}' = \Sigma_{i=1}^n r_i Z\gamma_{\sigma(i)}$.

PROOF. All parts are straightforward, except perhaps (iii):

$$\Sigma_{i=1}^n r_i = e'PJQ^{-\frac{1}{2}} = (0\ 0\ \cdots\ 0\ \sqrt{n})JQ^{-\frac{1}{2}} = 0 \quad . \qquad \text{QED}$$

LEMMA 4.5. $\mathfrak{Z}_2 = (n-2)!n\Sigma_{i=1}^n \gamma_i' Z'Z\gamma_i = (n-2)!n\Sigma_{i=1}^n \gamma_i' W'QW\gamma_i$.

PROOF. By Lemma 4.4 and (4.12), $\mathfrak{Z}_2 = \Sigma_\sigma^* \Sigma_{i,j=1}^n \gamma_{\sigma(i)}' Z' r_i' r_j Z\gamma_{\sigma(j)}$. This will be evaluated in two parts, one consisting of all terms with $i = j$, the other all terms with $i \neq j$.

$$\begin{aligned}
\Sigma_{i=1}^n \Sigma_\sigma^* \gamma_{\sigma(i)}' Z' r_i' r_i Z\gamma_{\sigma(i)} &= (n-1)!\Sigma_{i=1}^n \Sigma_{k=1}^n \gamma_k' Z' r_i' r_i Z\gamma_k \\
&= (n-1)!\Sigma_{k=1}^n \gamma_k' Z' \left[\Sigma_{i=1}^n r_i' r_i\right] Z\gamma_k \\
&= (n-1)!\Sigma_{k=1}^n \gamma_k' Z'Z\gamma_k \quad .
\end{aligned}$$

Observe that $\Sigma_{i\neq j} r_i' r_j = (\Sigma_i r_i)'(\Sigma_i r_i) - \Sigma_i r_i' r_i = -I$, so

$$\begin{aligned}
\Sigma_{i\neq j}\Sigma_\sigma^* \gamma_{\sigma(i)}' Z' r_i' r_j Z\gamma_{\sigma(j)} &= \Sigma_{i\neq j}(n-2)! \Big[\Big(\Sigma_{k=1}^n \gamma_k'\Big) Z' r_i' r_j Z\Big(\Sigma_{k=1}^n \gamma_k\Big) \\
&\qquad -\Sigma_{k=1}^n \gamma_k' Z' r_i' r_j Z\gamma_k\Big] \\
&= 0 - (n-2)!\Sigma_{k=1}^n \gamma_k' Z' \left[\Sigma_{i\neq j} r_i' r_j\right] Z\gamma_k \\
&= (n-2)!\Sigma_{k=1}^n \gamma_k' Z'Z\gamma_k \quad .
\end{aligned}$$

Combining these and using Lemma 4.4(i) completes the proof. QED

LEMMA 4.6. $v_1(T) = 0$ for all T, and $\frac{\partial}{\partial\Delta}\beta_\omega(\Delta)\big|_{\Delta=0} = 0$ for any (invariant) region ω .

PROOF. The second assertion follows from the first and (4.13). From (3.5),

$$\frac{\partial}{\partial\Delta} \log f_T(T|\Delta) = -\Delta\Sigma_{i=1}^n \gamma_i'\gamma_i + g'(\Delta)/g(\Delta) \quad . \tag{4.15}$$

This and Lemma 4.3 show that

$$\frac{\partial}{\partial\Delta} \log f_T(T|\Delta)\big|_{\Delta=0} \equiv v_1(T)/f_T(T|0)$$

is zero for all T . QED

LEMMA 4.7. $v_2(T) = 0$ for all T, and $\frac{\partial^2}{\partial\Delta^2}\beta_\omega(\Delta)\big|_{\Delta=0} = 0$ for any region ω .

PROOF. By (4.13) and Lemma 4.6, it suffices to show that

$$\frac{\partial^2}{\partial\Delta^2} \log f_T(T|\Delta)\big|_{\Delta=0} \equiv v_2(T)/f_T(T|0) - [v_1(T)/f_T(T|0)]^2 \tag{4.16}$$

is zero. Differentiating (4.15) gives

$$\frac{\partial^2}{\partial\Delta^2} \log f_T(T|\Delta) = -\Sigma_{i=1}^n \gamma_i'\gamma_i + g''(\Delta)/g(\Delta) - [g'(\Delta)/g(\Delta)]^2 \quad . \tag{4.17}$$

Lemma 4.3 gives $g'(0) = 0$, and Lemmas 4.2 and 4.5 give

$$g(0) = n!\int \mathrm{etr}(-\tfrac{1}{2}W'QW)|\det W|^{n-p-1}dW \quad ,$$

$$g''(0) = (n-2)!n\Sigma_{i=1}^n \int \gamma_i'W'QW\gamma_i \,\mathrm{etr}(-\tfrac{1}{2}W'QW)|\det W|^{n-p-1}dW \quad .$$

A change of variables from W to Z of (4.2) shows that

$$g(0) = n!(\det Q)^{-\frac{1}{2}(n-1)}\psi_0 \quad ,$$

$$g''(0) = (n-2)!n(\det Q)^{-\frac{1}{2}(n-1)}\Sigma_{i=1}^n \psi_i \quad ,$$

where

$$\psi_0 = \int \mathrm{etr}(-\tfrac{1}{2}Z'Z)(\det Z'Z)^{\frac{1}{2}(n-p-1)}dZ \quad ,$$

$$\psi_i = \int Y_i'Z'ZY_i\,\mathrm{etr}(-\tfrac{1}{2}Z'Z)(\det Z'Z)^{\frac{1}{2}(n-p-1)}dZ, \qquad i=1, \cdots, n \quad .$$

Integration is over all of p^2-dimensional Euclidean space.

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). For $q \geq p$, let $G\ell_{q,p} = \{Z(q \times p): \mathrm{rank}(Z) = p\}$, and define a measure μ_1 on $G\ell_{q,p}$ by

$$\mu_1(dZ) = dZ/(\det Z'Z)^{\frac{1}{2}q} \quad ,$$

where dZ is Lebesgue measure on $G\ell_{q,p}$. Let the function h: $G\ell_{q,p} \to S_p^+$ be defined by $h(Z) = Z'Z = S$, and the measure m on S_p^+ by

$$m(C) = \mu_1[h^{-1}(C)] \text{ for all measurable } C \subset S_p^+ \quad .$$

Then for any real-valued function g: $S_p^+ \to R$, integrable with respect to μ_1,

$$\int_{G\ell_{q,p}} g[h(Z)]\mu_1(dZ) = K\int_{S_p^+} g(S)m(dS) \quad . \tag{4.18}$$

To compute the measure m, note that it is invariant under the action of $G\ell_{p,p}$ on S_p^+ defined by $f_A(S) = ASA'$ for A in $G\ell_{p,p}$. This follows from (4.18) and the invariance of μ_1 under the action of $G\ell_{p,p}$ on $G\ell_{q,p}$ given by $f_A^*(Z) = ZA'$ for A in $G\ell_{p,p}$. The measure $dS/(\det S)^{\frac{1}{2}(p+1)}$ is invariant under $G\ell_{p,p}$. Since the measure on S_p^+ invariant under the action of $G\ell_{p,p}$ is unique up to a positive constant, it follows that, for all g,

$$\int_{G\ell_{q,p}} g[h(Z)]\mu_1(dZ) = K_Q\int_{S_p^+} g(S)\,\frac{dS}{(\det S)^{\frac{1}{2}(p+1)}} \quad ,$$

where the constant K_0, determined by μ_1 and h, is independent of g .

Apply this result to $\psi_i(i>0)$ and ψ_0 with $q=p$. Integrating over p^2-dimensional space will give the same result as integrating over $G\ell_{p,p}$, for their difference is a set of Lebesgue measure zero. Thus

$$\text{(4.19)} \qquad \begin{aligned} \psi_i &= K_0\int_{S_p^+} \gamma_i' S\gamma_i \operatorname{etr}(-\tfrac{1}{2}S)(\det S)^{\frac{1}{2}(n-1)}dS/(\det S)^{\frac{1}{2}(p+1)} \quad , \\ \psi_0 &= K_0\int_{S_p^+} \operatorname{etr}(-\tfrac{1}{2}S)(\det S)^{\frac{1}{2}(n-1)}dS/(\det S)^{\frac{1}{2}(p+1)} \quad . \end{aligned}$$

The first integral of (4.19) may be calculated by using a formula of Constantine (1963):

$$\text{(4.20)} \quad \int_{S_p^+} \operatorname{etr}(-RS)(\det S)^{t-\frac{1}{2}(p+1)} C_\varkappa(SU)dS = \Gamma_p(t,\varkappa)C_\varkappa(R^{-1}U)(\det R)^{-t} \quad ,$$

where $C_\varkappa$ is the zonal polynomial corresponding to the partition $\varkappa$. Take $R(p\times p) = \frac{1}{2}I$, $t = \frac{1}{2}(n-1)$, $U(p\times p) = \gamma_i\gamma_i'$, and $C_\varkappa$ the zonal polynomial corresponding to the partition (1) of the number 1 . The left-hand side of (4.20) equals the first integral of (4.19), for $C_1(SU) = \operatorname{tr}(SU) = \gamma_i' S\gamma_i$. It is routine to show that $C_1(R^{-1}U) = 2\gamma_i'\gamma_i$, $(\det R)^{-\frac{1}{2}(n-1)} = 2^{\frac{1}{2}(n-1)p}$, and (see Constantine, 1963, or Johnson and Kotz, 1972, p. 171) $\Gamma_p(t,\varkappa) = \frac{1}{2}(n-1)K_1$ where $K_1 = \pi^{\frac{1}{4}p(p-1)}\prod_{j=1}^p \Gamma\left(\frac{n-j}{2}\right)$. Combining these results yields

$$\psi_i = K_0K_1 2^{\frac{1}{2}(n-1)p}(n-1)\gamma_i'\gamma_i \quad .$$

The second integral of (4.19) is, up to a constant, a Wishart (I,p,n-1) density, so $\psi_0 = K_0K_1 2^{\frac{1}{2}(n-1)p}$. Thus $g''(0)/g(0) = \Sigma_{i=1}^n \gamma_i'\gamma_i$, which with Lemma 4.3 and (4.17) establishes that (4.16) equals zero. QED

LEMMA 4.8. $v_3(T) = 0$ for all T, and $\frac{\partial^3}{\partial\Delta^3}\beta_\omega(\Delta)\big|_{\Delta=0} = 0$ for any region ω.

PROOF. It suffices to show that

$$\frac{\partial^3}{\partial\Delta^3} \log f_T(T|\Delta)\big|_{\Delta=0} \equiv f^{-1}v_3(T) - 3f^{-2}v_1(T)v_2(T) + 2f^{-3}[v_1(T)]^3$$

is zero, where $f = f_T(T|0)$. Differentiating (4.17) and applying Lemma 4.3 establishes this. QED

THEOREM 4.3. For $j = 1, 2, 3$, $\frac{\partial^j}{\partial\Delta^j}\beta_\omega(\Delta)\big|_{\Delta=0} = 0$ for any region ω. $\frac{\partial^4}{\partial\Delta^4}\beta_\omega(\Delta)\big|_{\Delta=0}$ is maximized by a region of the form

$$\omega = \{T:\ g^{(4)}(0)/g(0) \geq k_0\} \quad ,$$

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

PROOF. The first assertion merely restates Lemmas 4.6, 4.7, and 4.8. The Generalized Neyman-Pearson Lemma shows that the region maximizing $\frac{\partial^4}{\partial\Delta^4}\beta_\omega(\Delta)\big|_{\Delta=0} = \int_\omega v_4(T)dT$ is

$$(4.21) \qquad \omega = \{T:\ v_4(T) \geq k_0v_0(T) + \cdots + k_3v_3(T)\} = \{T:\ v_4(T) \geq k_0v_0(T)\} \quad ,$$

since $v_1(T) = v_2(T) = v_3(T) = 0$ for all T.

It is routine to show that

$$\frac{\partial^4}{\partial\Delta^4} \log f_T(T|\Delta)\big|_{\Delta=0} = v_4(T)/v_0(T) \quad ,$$

and differentiating (4.17) twice and noting that $g'(0) = 0$ gives

$$\frac{\partial^4}{\partial\Delta^4} \log f_T(T|\Delta)\big|_{\Delta=0} = g^{(4)}(0)/g(0) - 3[g''(0)/g(0)]^2 \quad .$$

Thus (4.21) can be rewritten as

$$\omega = \{T:\ g^{(4)}(0)/g(0) - 3[g''(0)/g(0)]^2 \geq k_0\} \quad .$$

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

5. Evaluation of $g^{(4)}(0)$. A change of variables from W to Z in Lemma 4.2 yields

$$(5.1)\qquad g^{(4)}(0) = (\det Q)^{-\frac{1}{2}(n-1)} \int \mathfrak{z}_4 \operatorname{etr}(-\tfrac{1}{2}Z'Z)|\det Z|^{n-p-1} dZ \quad .$$

A useful expression for $\mathfrak{z}_4$ follows.

THEOREM 5.1.

$$\mathfrak{z}_4 = (n-4)![(n^3+n^2)\mathscr{A}_1 + (3n^2-9n+3)\mathscr{A}_2 - (3n^2-3n-6)\mathscr{A}_3 - (3n^2-3n)\mathscr{A}_4 + 6\mathscr{A}_5] \quad ,$$

where

$$\mathscr{A}_1 = \Sigma_{i,j=1}^n (r_i Z\gamma_j)^4, \quad \mathscr{A}_2 = \left(\Sigma_{i=1}^n \gamma_i' Z'Z\gamma_i\right)^2, \quad \mathscr{A}_3 = \Sigma_{i=1}^n (\gamma_i' Z'Z\gamma_i)^2,$$

$$\mathscr{A}_4 = \Sigma_{i=1}^n \left[\Sigma_{j=1}^n (r_i Z\gamma_j)^2\right]^2, \quad \text{and} \quad \mathscr{A}_5 = \Sigma_{\substack{i,j=1 \\ i\neq j}}^n (\gamma_i' Z'Z\gamma_j)^2 \quad .$$

PROOF. $\mathfrak{z}_4 = \Sigma_\sigma^* \left[\Sigma_{i=1}^n r_i Z\gamma_{\sigma(i)}\right]^4$ by Lemma 4.4(v). It is an exercise to verify the identity

$$(\Sigma x_i)^4 = \Sigma_i x_i^4 + 4\Sigma_{ij\neq} x_i^3 x_j + 3\Sigma_{ij\neq} x_i^2 x_j^2 + 6\Sigma_{ijk\neq} x_i^2 x_j x_k + \Sigma_{ijkh\neq} x_i x_j x_k x_h \quad ,$$

where each sum on the right-hand side is taken over all sets of distinct subscripts. Letting $x_i = x_{\sigma,i} = r_i Z\gamma_{\sigma(i)}$ and summing this identity over all permutations σ gives

$$(5.2)\qquad \mathfrak{z}_4 = \Sigma_\sigma^*[\Sigma_i x_{\sigma,i}]^4 = \Sigma_\sigma^*\{\Sigma_i x_{\sigma,i}^4 + 4\Sigma_{ij\neq} x_{\sigma,i}^3 x_{\sigma,j} + 3\Sigma_{ij\neq} x_{\sigma,i}^2 x_{\sigma,j}^2 + \cdots\} \quad .$$

Each sum on the right can be expressed in terms of $\mathscr{A}_1$ to $\mathscr{A}_5$ by repeated use of parts (iii) and (iv) of Lemma 4.4.

(i) $\Sigma_\sigma^* \Sigma_i x_{\sigma,i}^4 = (n-1)!\Sigma_{i,i'}(r_i Z\gamma_{i'})^4 = (n-1)!\mathscr{A}_1$.

(ii) $\Sigma^*_\sigma \Sigma_{ij\neq} x^3_{\sigma,i} x_{\sigma,j} = (n-2)!\Sigma_{ij\neq,i'j'\neq}(r_i Z\gamma_{i'})^3 (r_j Z\gamma_{j'})$

$$= (n-2)!\Sigma_{i,i'}(r_i Z\gamma_{i'})^3 (\Sigma_{j\neq i} r_j) Z(\Sigma_{j'\neq i'}\gamma_{j'})$$

$$= (n-2)!\mathscr{A}_1 \quad .$$

(iii) $\Sigma^*_\sigma \Sigma_{ij\neq} x^2_{\sigma,i} x^2_{\sigma,j} = (n-2)!\Sigma_{ij\neq,i'j'\neq}(r_i Z\gamma_{i'})^2 (r_j Z\gamma_{j'})^2$

$$= (n-2)!\Sigma_{i'j'\neq,i}(r_i Z\gamma_{i'})^2 \gamma'_{j'} Z'(I - r'_i r_i) Z\gamma_{j'}$$

$$= (n-2)!\left\{\Sigma_{i'j'\neq}\gamma'_{i'}Z'Z\gamma_{i'}\gamma'_{j'}Z'Z\gamma_{j'} - \Sigma_i\left[\Sigma_{i'}(r_i Z\gamma_{i'})^2\right]^2 + \Sigma_{i,i'}(r_i Z\gamma_{i'})^4\right\}$$

$$= (n-2)!\left\{(\mathscr{A}_2 - \mathscr{A}_3) - \mathscr{A}_4 + \mathscr{A}_1\right\} \quad .$$

The derivation of the remaining two sums is technically involved but similar. Details may be found in Schwager (1979). Substituting these results into (5.2) produces

$$\mathfrak{I}_4 = (n-1)!\mathscr{A}_1 + 4\cdot(n-2)!\mathscr{A}_1 + 3\cdot(n-2)!(\mathscr{A}_1 + \mathscr{A}_2 - \mathscr{A}_3 - \mathscr{A}_4)$$
$$+ 6\cdot(n-3)!(4\mathscr{A}_1 + \mathscr{A}_2 - 2\mathscr{A}_3 - 2\mathscr{A}_4)$$
$$+ (n-4)!(36\mathscr{A}_1 + 3\mathscr{A}_2 - 12\mathscr{A}_3 - 18\mathscr{A}_4 + 6\mathscr{A}_5) \quad .$$

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

COROLLARY 5.1.

$$g^{(4)}(0) = (n-4)!(\det Q)^{-\frac{1}{2}(n-1)}\int[(n^3 + n^2)\mathscr{A}_1 + (3n^2 - 9n + 3)\mathscr{A}_2$$
$$- (3n^2 - 3n - 6)\mathscr{A}_3 - (3n^2 - 3n)\mathscr{A}_4 + 6\mathscr{A}_5]\mathrm{etr}(-\tfrac{1}{2}Z'Z)|\det Z|^{n-p-1}dZ \quad ,$$

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

Only two of the integrals $\int\mathscr{A}_i \mathrm{etr}(-\frac{1}{2}Z'Z)|\det Z|^{n-p-1}dZ$, $i = 1, \cdots, 5$, must be calculated, for conditional on A, the integrals with leading terms $\mathscr{A}_2$, $\mathscr{A}_3$, and $\mathscr{A}_5$ are constants. Their integrands are functions only of the variable of integration, Z, and the γ_i's, which are determined by A . The integral

$$\int[(3n^2 - 9n + 3)\mathscr{S}_2 - (3n^2 - 3n - 6)\mathscr{S}_3 + 6\mathscr{S}_5]\operatorname{etr}(-\tfrac{1}{2}Z'Z)|\det Z|^{n-p-1}dZ$$

can therefore be denoted by k_1, a constant depending on A, n, and p, but not on the data Y . 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

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

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,

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

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. QED

COROLLARY 5.2.

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

LEMMA 5.2.

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

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

$$\Phi = \int[(x_{11} \cdots x_{1p})P_1]^4 \operatorname{etr}(-\tfrac{1}{2}X'X)|\det X|^{n-p-1}dX \quad .$$

Letting P_1 equal $(2^{-\frac{1}{2}} \pm 2^{-\frac{1}{2}}\ 0 \cdots 0)$ and $(0\ 1\ 0 \cdots 0)$ gives

$$\Phi = \int [2^{-\frac{1}{2}}(x_{11} \pm x_{12})]^4 \operatorname{etr}(-\tfrac{1}{2}X'X)|\det X|^{n-p-1} dX$$
$$= \int x_{12}^4 \operatorname{etr}(-\tfrac{1}{2}X'X)|\det X|^{n-p-1} dX \quad .$$

Now multiply both sides of the identity

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

by $\operatorname{etr}(-\frac{1}{2}Z'Z)|\det Z|^{n-p-1}$ and integrate with respect to dZ . QED

THEOREM 5.2.

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

PROOF. For any i, by simultaneous diagonalization (see Press, 1972, p. 37), there exists an orthogonal $p \times p$ matrix U such that

$$U(r_i'r_i)U' = \operatorname{diag}(\|r_i\|^2, 0, 0, \cdots, 0) \quad .$$

Define a $p \times p$ matrix variable X by $X = UZ$, so that

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

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

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

it follows that

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

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

$$(5.3) \qquad \int \mathscr{A}_4 \operatorname{etr}(-\tfrac{1}{2}Z'Z)|\det Z|^{n-p-1} dZ = \left[\Sigma_{i=1}^n \|r_i\|^4\right] \Pi \quad .$$

To evaluate Π, we note two relations involving the eigenvalues $\lambda_1, \cdots, \lambda_p$ of C:

$$(5.4)\qquad \begin{aligned} \Sigma_{i=1}^{p}\lambda_i &= \operatorname{tr} C &&= \Sigma_{j=1}^{n}\|\gamma_j\|^2 \quad , \\ \Sigma_{i=1}^{p}\lambda_i^2 &= \operatorname{tr}(C^2) &&= \Sigma_{j,k=1}^{n}(\gamma_j'\gamma_k)^2 \quad . \end{aligned}$$

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

$$\begin{aligned} \Pi &= \int \left[\Sigma_{i=1}^{p}\lambda_i y_{1i}^2\right]^2 \operatorname{etr}(-\tfrac{1}{2}Y'Y)\,|\det Y|^{n-p-1}dY \\ &= \left[\Sigma_{i=1}^{p}\lambda_i^2\right]\Phi + \left[\Sigma_{\substack{i,j=1 \\ i\neq j}}^{p}\lambda_i\lambda_j\right]\tfrac{1}{3}\Phi \\ &= \tfrac{1}{3}\Phi\left[2\Sigma_{i=1}^{p}\lambda_i^2 + \left(\Sigma_{i=1}^{p}\lambda_i\right)^2\right] \quad . \end{aligned}$$

Substitute this and (5.4) into (5.3) to complete the proof. QED

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

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

$$g^{(4)}(0) = (n-4)!(\det Q)^{-\frac{1}{2}(n-1)}\left[\Phi L \Sigma_{i=1}^{n}\|r_i\|^4 + k_1\right] \quad ,$$

where

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

The constant k_1 depends on A, n, and p .

PROOF. It was observed following Corollary 5.1 that the three integrals that do not depend on the data result in a constant k_1 . Substituting this and the results of Corollary 5.2 and Theorem 5.2 into the formula of Corollary 5.1 for $g^{(4)}(0)$ establishes the theorem. QED

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

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

LEMMA 6.1.
$$b_{2,p} = n\Sigma_{i=1}^{n}\|r_i\|^4 \quad .$$

PROOF. It follows from (3.1) and (3.4) that $P'M\widetilde{Y} = JW$, and thus $\widetilde{Y} - e\bar{Y} = M\widetilde{Y} = PJW$. Since $S = (\widetilde{Y} - e\bar{Y})'(\widetilde{Y} - e\bar{Y}) = W'QW$, it follows that

$$[Y_{(i)} - \bar{Y}]S^{-1}[Y_{(i)} - \bar{Y}]' = P_iJQ^{-1}J'P_i' = \|r_i\|^2 \quad .$$

Substituting into the formula defining $b_{2,p}$ proves the lemma. QED

LEMMA 6.2. The critical region maximizing $(\partial^4/\partial\Delta^4)\beta_\omega(\Delta)\big|_{\Delta=0}$ is specified by $L\Sigma_{i=1}^{n}\|r_i\|^4 \geq k_0'$, where L and k_0' are functions of A .

PROOF. Substituting the result of Theorem 5.3 and the value of g(0) from Section 4 into Theorem 4.3 gives

$$(6.1) \qquad (n-4)![\Phi L\Sigma_i\|r_i\|^4 + k_1]/k_2 \geq k_0 \quad ,$$

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 completes the proof. QED

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 of Lemma 6.2. If L is positive, this region can be specified by

$$b_{2,p} = n\Sigma_i \|r_i\|^4 \geq nk_0'/L = K \quad ;$$

if L is negative, by

$$b_{2,p} = n\Sigma_i \|r_i\|^4 \leq nk_0'/L = K' \quad . \qquad \text{QED}$$

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 $(3 - \sqrt{3})/6 = 21.13 \cdots \%$. 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 which 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

$$(6.2) \qquad (n^3 + n^2)\Sigma_i \|\gamma_i\|^4 > (n^2 - n)(2F + 1)(\Sigma_i \|\gamma_i\|^2)^2 \quad ,$$

where

$$F = \Sigma_{i,j} (\gamma_i' \gamma_j)^2 / (\Sigma_i \|\gamma_i\|^2)^2 \quad .$$

Observing that F is nonnegative and summing the Cauchy-Schwarz inequality

$(\gamma_i'\gamma_j)^2 \le \|\gamma_i\|^2\|\gamma_j\|^2$ over all i and j, one sees that $0 \le F \le 1$. Consequently, it suffices for (6.2) to prove that

(6.3) $$(n^3 + n^2)\Sigma_i\|\gamma_i\|^4 > 3(n^2 - n)(\Sigma_i\|\gamma_i\|^2)^2 \quad .$$

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,

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

Also, taking the squared norm of each side in the identity

$$(n-m)\bar{a}' = \Sigma_{i=1}^m\gamma_i$$

and repeatedly applying the inequality $\gamma_i'\gamma_j + \gamma_j'\gamma_i \le \gamma_i'\gamma_i + \gamma_j'\gamma_j$ yields

(6.5) $$(n-m)^2\|\bar{a}\|^2 \le m\Sigma_{i=1}^m\|\gamma_i\|^2 \quad .$$

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

(6.6) $$G(x_1, \cdots, x_m, y) = \frac{\left[\Sigma_{i=1}^m x_i + (n-m)y\right]^2}{\Sigma_{i=1}^m x_i^2 + (n-m)y^2} < \frac{n^2+n}{3(n-1)} \quad .$$

The condition (6.5), which A must satisfy, may be written as

$$(n-m)^2 y \le m\Sigma_{i=1}^m x_i \quad .$$

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

(6.7) $$\left\{(x_1, \cdots, x_m, y) : x_1, \cdots, x_m, y \ge 0; (n-m)^2 y \le m\Sigma_{i=1}^m x_i\right\} \quad ,$$

begin by observing that G is increased by equalizing the x_i; that is, for $\bar{x} = m^{-1}\Sigma_{i=1}^{m} x_i$,

$$G(x_1, \cdots, x_m, y) \leq G(\bar{x}, \cdots, \bar{x}, y) \quad ,$$

with equality only if all x_i are equal. To find the maximum value of G for fixed m and n, consider G as a function of two variables,

$$G(\bar{x}, y) = [m\bar{x} + (n-m)y]^2/[m\bar{x}^2 + (n-m)y^2] \quad ,$$

subject to the condition $0 \leq (n-m)^2 y \leq m^2\bar{x}$. 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 at $u = m^2/(n-m)^2$, where

$$G(u) = n^2 m(n-m)/[m^3 + (n-m)^3] \quad .$$

It follows that a sufficient condition for (6.6) to hold on (6.7) is that

$$(6.8) \qquad n^2 m(n-m)/[m^3 + (n-m)^3] < (n^2+n)/3(n-1) \quad .$$

Let $s = m/n$; (6.8) holds whenever

$$s(1-s)/[s^3 + (1-s)^3] \leq \tfrac{1}{3} \quad ,$$

or equivalently whenever $6s^2 - 6s + 1 \geq 0$.

Solving this quadratic inequality, and restricting attention to the interval $(0, \frac{1}{2})$ as required by the model, gives

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

THEOREM 6.3. If $a_i = 0$ for $i = m+1, \cdots, n$, $\Sigma_{i=1}^{m} a_i = 0$, and $m/n \leq \frac{1}{3}$, then $L > 0$, and the test which 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, (6.6) becomes

$$(6.9) \qquad G(X) = X'EX/X'X < (n^2+n)/3(n-1) \quad .$$

The maximum of $X'EX/X'X$ is m, the largest eigenvalue of E, so (6.9) holds for all X when

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

for which $m/n \leq \frac{1}{3}$ is sufficient. QED

If the general outlier problem is assumed to have a fraction of outliers no greater than $21.13 \cdots \%$, 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 $33\frac{1}{3}\%$. 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.

REFERENCES

[1] Anscombe, F. J. (1960). Rejection of outliers. Technometrics 2, 123-147.

[2] Anscombe, F. J. and Tukey, J. W. (1963). The examination and analysis of residuals. Technometrics 5, 141-160.

[3] Barnett, V. and Lewis, T. (1978). Outliers in Statistical Data. John Wiley and Sons, New York.

[4] Butler, R. W. (1977). Cross-validatory predictive intervals, the Bayes character of subset selection techniques, and robust adaptive prediction. Ph.D. thesis, The Univ. of Michigan.

[5] Constantine, A. G. (1963). Noncentral distribution problems in multivariate analysis. Ann. Math. Statist. 34, 1270-1285.

[6] David, H. A. (1970). Order Statistics. John Wiley and Sons, New York.

[7] Doornbos, R. (1966). Slippage Tests. Mathematical Centre Tracts 15, Mathematisch Centrum, Amsterdam.

[8] Eaton, M. L. (1972). Multivariate Statistical Analysis. Inst. Math. Statist., Univ. Copenhagen, Copenhagen, Denmark.

[9] Ferguson, T. S. (1961). On the rejection of outliers. Proc. Fourth Berkeley Symposium Math. Statist. Prob. I, 253-287.

[10] Ferguson, T. S. (1967). Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York.

[11] Gnanadesikan, R. (1977). Methods for Statistical Data Analysis of Multivariate Observations. John Wiley and Sons, New York.

[12] Healy, M. J. R. (1968). Multivariate normal plotting. Appl. Statist. 17, 157-161.

[13] Johnson, N. L. and Kotz, S. (1972). Distributions in Statistics: Continuous Multivariate Distributions. John Wiley and Sons, New York.

[14] Karlin, S. and Truax, D. R. (1960). Slippage problems. Ann. Math. Statist. 31, 296-324.

[15] Kruskal, W. H. (1960). Some remarks on wild observations. Technometrics 2, 1-3.

[16] Layard, M. W. J. (1974). A Monte Carlo comparison of tests for equality of covariance matrices. Biometrika 61, 461-465.

[17] Loeve, M. (1977). Probability Theory I, 4th Edition. Springer-Verlag, New York.

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

[19] Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. Sankhyā B 36, 115-128.

[20] Mardia, K. V. (1975). Assessment of multinormality and the robustness of Hotelling's T^2 test. Appl. Statist. 24, 163-171.

[21] Press, S. J. (1972). Applied Multivariate Analysis. Holt, Rinehart and Winston, Inc., New York.

[22] Rohlf, F. J. (1975). Generalization of the gap test for the detection of multivariate outliers. Biometrics 31, 93-101.

[23] Schwager, S. J. (1979). Detection of multivariate outliers. Ph.D. thesis, Yale Univ.

[24] 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.

[25] Wilks, S. S. (1963). Multivariate statistical outliers. Sankhyā A 25, 406-427.