ON THE REJECTION OF OUTLIERS

THOMAS S. FERGUSON
UNIVERSITY OF CALIFORNIA, LOS ANGELES AND PRINCETON UNIVERSITY

1. Introduction and summary

The general problem treated in this paper is a very old and common one. In its simplest form it may be stated as follows. In a sample of moderate size taken from a certain population, it appears that one or two values are surprisingly far away from the main group. The experimenter is tempted to throw away the apparently erroneous values, and not because he is certain that the values are spurious. On the contrary, he will undoubtedly admit that even if the population has a normal distribution there is a positive although extremely small probability that such values will occur in an experiment. It is rather because he feels that other explanations are more plausible, and that the loss in the accuracy of the experiment caused by throwing away a couple of good values is small compared to the loss caused by keeping even one bad value. The problem, then, is to introduce some degree of objectivity into the rejection of the outlying observations.

There is no need to give here a historical outline of progress in the subject. Good accounts of the historical aspect, interesting because this subject was one of the first problems to receive a statistical treatment, may be found in recent papers by Grubbs [4], Murphy [9], and Anscombe [1]. We shall mention here only those papers which directly concern us, with particular emphasis placed on the last ten years.

Two mathematical models have been proposed, implicitly by Grubbs and explicitly by Dixon [3], to give a structure to the outlier problem. In both models it is assumed that a sample of n observations, to be denoted by $X_1, \cdots, X_n$ has been drawn from a normal population with possibly unknown mean and variance. A few of these values may have been spuriously changed. In *model A*, this change is hypothesized as a shift in the mean, and in *model B*, as an increase in the variance. The precise formulations will be made clear in the particular problems we shall consider.

This paper is mainly concerned with the derivation, found in section 2, of the locally best tests for the existence of spurious observations in several situations for both models A and B. The tests suggested here by virtue of their local optimality are based on the sample coefficient of skewness for one-sided alternatives, and on the sample coefficient of kurtosis for two-sided alternatives. Many spurious observations are allowed under the alternative hypotheses; even in the most

Prepared in connection with research sponsored by the Office of Naval Research Contract Nonr-1858(05), Project ONR 042-023.

stringent case, 21% of the observations are allowed to be spurious. It is seen in section 2.4 that these tests have a very strong locally optimum property.

A very satisfactory optimum property of the rejection rule based on the maximum Studentized deviation from the mean has been established independently by Paulson [11], and Murphy [9], and later by A. Kudo [7]. It was shown, essentially, that under model A assumptions this rule will maximize the probability of rejecting the spurious observation when there is only one spurious observation. In section 3, this rule is seen to have the same optimum property under model B assumptions. The proof is carried out in many dimensions, as was done for model A by Karlin and Truax [6].

It is the purpose of section 4 to determine how well the optimum property discovered by Paulson and Murphy will extend to the situation in which the observations are from an experiment with a more complicated design, for the rejection rule based on the maximum Studentized residual. This rejection rule has been discussed in recent papers by Anscombe [1] and Daniel [2]. It will be seen that if one restricts oneself to invariant rules, the proposed rule is admissible.

The last section contains the results of some sampling experiments performed with the view of giving a better picture of the relative performance of certain of these rejection rules.

2. Locally best tests for the existence of outliers

We shall deal in this section with the derivation of (1) the one-sided locally best invariant test, (2) the locally best unbiased invariant test, (3) the locally best invariant test when the mean is known, and (4) the locally best invariant test under model B.

2.1 *Model* A, *mean unknown.* First, the locally best invariant tests for the existence of outliers under model A, mean unknown, will be derived. In this context, model A specifies that $X_1, X_2, \cdots, X_n$, *are independent random variables having normal distributions with a common unknown variance* $\sigma^2 > 0$. *We assume that there are known real numbers* $a_1, a_2, \cdots, a_n$, *and unknown parameters,* μ, Δ, *and an unknown permutation* $(\nu_1, \nu_2, \cdots, \nu_n)$ *of the first n positive integers, such that* $EX_i = \mu + \sigma\Delta a_{\nu_i}$ *for* $i = 1, \cdots n$. *The variance of the* a_i *is assumed to be positive.* It is usual in outlier problems to assume that most of the a_i are zero. We will make certain restrictions of this sort later in dealing with specific problems. In particular it will be shown that the locally best invariant tests derived here are locally best uniformly over certain sets of (a_i) satisfying natural conditions.

We shall denote the null hypothesis always by H_0 and the alternatives by $\bar{H}$ or H_1, H_2, and so on. In this case we have

$$\begin{aligned} &H_0: &\Delta = 0, \\ &\bar{H}: &\Delta \neq 0. \end{aligned} \tag{2.1}$$

We shall also consider the one-sided tests whose alternatives are

$$H_1: \quad \Delta > 0. \tag{2.2}$$

The problem is obviously invariant under the following three transformations: (1) any permutation of the subscripts of $X_1, X_2, \cdots, X_n$; (2) the addition of a constant to each X_i; (3) multiplication of each X_i by a positive constant.

We would like to consider only those rejection criteria which also are invariant under these three transformations. Since the rule is to be invariant under (1), we need consider only functions of the order statistics $X_{(1)}$, $X_{(2)}, \cdots, X_{(n)}$. Since the rule is to be invariant under (2), we need consider only functions of the differences $X_{(2)} - X_{(1)}, X_{(3)} - X_{(1)}, \cdots, X_{(n)} - X_{(1)}$. Since the rule is to be invariant under (3), we need consider only functions of the ratios, $(X_{(2)} - X_{(1)})/(X_{(n)} - X_{(1)}), \cdots, (X_{(n-1)} - X_{(1)})/(X_{(n)} - X_{(1)})$. We shall now proceed to derive the distribution of these ratios. Since this distribution does not depend on μ or σ, we shall from the beginning take $\mu = 0$, and $\sigma = 1$ in order to simplify the algebra.

The joint density of $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ under model A, with $\mu = 0$ and $\sigma = 1$, is

$$(2.3)\quad f_{X_{(1)},\cdots,X_{(n)}}(x_1, \cdots, x_n) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left[-\frac{1}{2}\sum x_i^2 - \frac{1}{2}\Delta^2 \sum a_i^2\right] \sum{}^* \exp\left[\Delta \sum_{j=1}^{n} x_{\nu_j} a_j\right],$$

where $x_1 < x_2 < \cdots < x_n$ and where $\sum^*$ denotes the summation over all permutations, $(\nu_1, \cdots, \nu_n)$ of the first n positive integers. Now we make the transformation,

$$(2.4)\quad \begin{aligned} W &= X_{(1)}, \\ Z_i &= X_{(i)} - X_{(1)} \qquad \text{for } i = 2, \cdots, n, \\ X_{(1)} &= W, \\ X_{(i)} &= W + Z_i \qquad \text{for } i = 2, \cdots, n. \end{aligned}$$

The Jacobian of this transformation is one. After making the transformation, we integrate W from $-\infty$ to ∞, to find the joint density of $Z_2, \cdots, Z_n$. In the following formulas the dummy variable z_1 denotes zero.

$$(2.5)\quad \begin{aligned} f_{Z_2,\cdots,Z_n}(z_2, \cdots, z_n) &= c \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\sum_{i=1}^{n} (w + z_i)^2 - \frac{1}{2}\Delta^2 \sum a_j^2\right] \sum{}^* \exp\left[\Delta \sum (w + z_{\nu_j}) a_j\right] dw \\ &= c \exp\left[-\frac{1}{2}\sum (z_i - \bar{z})^2 - \frac{1}{2}\Delta^2 \sum (a_j - \bar{a})^2\right] \sum{}^* \exp\left[\Delta \sum (z_{\nu_j} - \bar{z}) a_j\right], \end{aligned}$$

where $0 = z_1 < z_2 < \cdots < z_n$, where $\bar{z} = (1/n) \sum z_i$ and $\bar{a} = (1/n) \sum a_i$, and where c will always denote the constant necessary to make the formula have total probability one.

Now we shall make the transformation

$$(2.6)\quad \begin{aligned} Y_i &= \frac{Z_i}{Z_n}, \quad i = 2, \cdots, n-1, \qquad & Z_i &= Y_i V, \quad i = 2, \cdots, n-1, \\ V &= Z_n, & Z_n &= V. \end{aligned}$$

The Jacobian of this transformation is V^{n-2}. After making this transformation we integrate V from 0 to ∞, to find the joint density of $Y_2, \cdots, Y_{n-1}$. In the following formulas the dummy variables y_1 and y_n will denote zero and one respectively, $\bar{y}$ will denote $n^{-1}\sum_1^n y_i$, and s^2 will denote $n^{-1}\sum_1^n (y_i - \bar{y})^2$.

$$(2.7)\qquad f_{Y_2,\cdots,Y_{n-1}}(y_2, \cdots, y_{n-1})$$
$$= c\exp\left[-\frac{\Delta^2}{2}\sum (a_j - \bar{a})^2\right]\sum{}^* \int_0^\infty \exp\left[-\frac{1}{2}v^2 ns^2 + v\Delta\sum (y_{\nu_i} - \bar{y})a_j\right]v^{n-2}\,dv$$
$$= c\exp\left[-\frac{\Delta^2}{2}\sum (a_j - \bar{a})^2\right]s^{-(n-1)}g(\Delta),$$

where

$$(2.8)\qquad g(\Delta) = \sum{}^*\int_0^\infty \exp\left[-\frac{t^2}{2} + t\Delta\sum u_{\nu_j}a_j\right]t^{n-2}dt$$

where $0 < y_2 < \cdots < y_{n-1} < 1$, where $u_i = (y_i - \bar{y}) / \sqrt{n}\,s$, and where we have made the change of variable $t = \sqrt{n}\,sv$.

This distribution depends on only one parameter, Δ, so that we may apply the method of Neyman and Pearson [11] in deriving the locally best test of the hypothesis $H_0: \Delta = 0$ against either of the alternatives $\overline{H}: \Delta \neq 0$ or $H_1: \Delta > 0$. In using this method it is necessary for us to pass a certain number of derivatives under the integral sign. To justify this we shall use the criterion (see Loève [8], p. 126) that whenever for some $\epsilon > 0$

$$(2.9)\qquad \left|\frac{\partial}{\partial\theta}f(x, \theta)\right| < h(x)$$

for all $|\theta - \theta_0| < \epsilon$, where $h(x)$ is integrable, then

$$(2.10)\qquad \frac{\partial}{\partial\theta}\int f(x, \theta)\,dx\Big|_{\theta=\theta_0} = \int \frac{\partial f(x, \theta)}{\partial\theta}\Big|_{\theta=\theta_0}dx.$$

We first apply this criterion to the integral in the expression for the density of $Y_2, \cdots, Y_{n-1}$. Here, for any value of $|\Delta| < B$,

$$(2.11)\qquad \left|\frac{\partial}{\partial\Delta}\exp\left[-\frac{t^2}{2} + t\Delta\sum u_{\nu_j}a_j\right]t^{n-2}\right|$$
$$= \exp\left[-\frac{t^2}{2} + t\Delta\sum u_{\nu_j}a_j\right]t^{n-1}\left|\sum u_{\nu_j}a_j\right|$$
$$\leqq \exp\left[-\frac{t^2}{2} + t\left|B\sum u_{\nu_j}a_j\right|\right]t^{n-1}\left|\sum u_{\nu_j}a_j\right|,$$

which is an integrable function. It is apparent that derivatives of any order may be taken under the integral sign at all finite values of Δ.

Next we apply this criterion to the power function $\beta_\omega(\Delta)$. If ω is any rejection region, that is, a measurable subset of R^{n-2}, the power function corresponding to this region is defined by the formula

$$\beta_\omega(\Delta) = P[(Y_2, \cdots, Y_{n-1}) \in \omega|\Delta] = \int_\omega f_{Y_2,\cdots,Y_{n-1}}(y_2, \cdots, y_{n-1}|\Delta)\, dy_2 \cdots dy_{n-1}. \tag{2.12}$$

We shall need to take several derivatives of $\beta_\omega(\Delta)$ at $\Delta = 0$, and in order to apply the Neyman-Pearson theory we shall need to be able to evaluate these derivatives by passing them under the integral sign in (2.12). For all values of Δ such that $|\Delta| < B$, we have

$$\begin{aligned}\left|\frac{\partial}{\partial\Delta} f_{Y_2,\cdots,Y_{n-1}}(y_2, \cdots, y_{n-1}|\Delta)\right| &\leqq cs^{-(n-1)} \exp\left[-\frac{\Delta^2}{2}\sum (a_j - \bar{a})^2\right]\Bigg|\Delta \sum (a_j - \bar{a})^2 g(\Delta) \\ &\quad + \sum{}^*\sum u_{\nu_j}a_j \int_0^\infty \exp\left[-\frac{t^2}{2} + t\,\Delta \sum u_{\nu_j}a_j\right] t^{n-1}\,dt\Bigg| \\ &\leqq cs^{-(n-1)} \exp\left[-\frac{\Delta^2}{2}\sum (a_j - \bar{a})^2\right] \\ &\quad \Bigg\{B \sum (a_j - \bar{a})^2 \sum{}^* \int_0^\infty \exp\left[-\frac{t^2}{2} + tB|\sum u_{\nu_j}a_j|\right] t^{n-2}\,dt \\ &\quad + \sum{}^* |\sum u_{\nu_j}a_j| \int_0^\infty \exp\left[-\frac{t^2}{2} + tB|\sum u_{\nu_j}a_j|\right] t^{n-1}\,dt\Bigg\},\end{aligned} \tag{2.13}$$

which is integrable $dy_2 \cdots dy_{n-1}$ over $0 < y_2 < \cdots < y_{n-1}$. ($s^2$ is bounded from below by $(2n)^{-1}$.) Again it is easy to see that derivatives of any order may be taken under the integral sign (2.12) at all finite values of Δ and for all measurable sets ω.

We are interested in the behavior of $\beta_\omega(\Delta)$ for values of Δ in a neighborhood of $\Delta = 0$. We will first show that the first two derivatives of $\beta_\omega(\Delta)$ vanish at the origin for every (invariant) region ω. Since the derivatives may be placed under the integral sign, it is sufficient to show that

$$\left.\frac{\partial^i}{\partial\Delta^i} \log f_{Y_2,\cdots,Y_{n-1}}(y_2, \cdots, y_{n-1}|\Delta)\right|_{\Delta=0} = 0 \tag{2.14}$$

identically in $y_2, \cdots, y_{n-1}$, and for $i = 1$ and 2. Note first that

$$\sum_{i=1}^n u_i = 0, \qquad \sum_{i=1}^n u_i^2 = 1, \tag{2.15}$$

and that

$$\frac{\partial}{\partial\Delta} \log f_{Y_2,\cdots,Y_{n-1}}(y_2, \cdots, y_{n-1}|\Delta) = -\Delta \sum (a_j - \bar{a})^2 + \frac{g'(\Delta)}{g(\Delta)}. \tag{2.16}$$

Let

$$C_n = \int_0^\infty \exp\left[-\frac{1}{2}t^2\right] t^n dt = 2^{(n-1)/2}\,\Gamma\left(\frac{n+1}{2}\right) \tag{2.17}$$

so that

$$(2.18)\qquad \begin{aligned} g(0) &= C_{n-2}\, n!,\\ g'(0) &= C_{n-1} \sum{}^* \sum_{j=1}^{n} u_{\nu_j} a_j = C_{n-1} \sum_{j=1}^{n} a_j\, (n-1)! \sum_{\nu=1}^{n} u_\nu = 0, \end{aligned}$$

using (2.15). Hence we see from (2.16) that (2.14) is true for $i = 1$. We take another derivative of (2.16).

$$(2.19)\qquad \frac{\partial^2}{\partial\Delta^2} \log f = - \sum (a_j - \bar{a})^2 + \frac{g(\Delta)g''(\Delta) - g'(\Delta)^2}{g(\Delta)^2},$$

$$(2.20)\qquad \begin{aligned} g''(0) &= C_n \sum{}^* \Big(\sum_{j=1}^{n} u_{\nu_j} a_j\Big)^2 = C_n \sum_{j=1}^{n} \sum_{k=1}^{n} a_j\, a_k \sum{}^* u_{\nu_j} u_{\nu_k}\\ &= C_n\, (n-2)!\, \Big[(n-1) \sum_{j=1}^{n} a_j^2 - \sum_{j \neq k} a_j\, a_k\Big], \end{aligned}$$

$$(2.21)\qquad \begin{aligned} \frac{g''(0)}{g(0)} &= \left(\frac{C_n}{C_{n-2}}\right) \frac{1}{n(n-1)} \left[n \sum a_j^2 - \Big(\sum a_j\Big)^2\right]\\ &= \sum (a_j - \bar{a})^2. \end{aligned}$$

Hence it follows from (2.21) and (2.19) that (2.14) is true for $i = 2$.

We are thus confronted with the fact that if we wish to distinguish between two rejection regions on the basis of their behavior locally at $\Delta = 0$, we must look to derivatives of order higher than two. It is easy to see, following the Neyman-Pearson theory, that the locally best rejection region for testing H_0 against H_1 consists of those points for which

$$(2.22)\qquad \frac{\partial^3}{\partial\Delta^3} \log f_{Y_2,\cdots,Y_{n-1}}(y_2, \cdots, y_{n-1}|\Delta)\Big|_{\Delta=0} \geqq K_1,$$

where K_1 is chosen so that the probability of rejecting H_0 when it is true is a fixed value α chosen in advance. One easily finds that

$$(2.23)\qquad \frac{\partial^3}{\partial\Delta^3} \log f_{Y_2,\cdots,Y_{n-1}}(y_2, \cdots, y_{n-1}|\Delta)\Big|_{\Delta=0} = \frac{g'''(0)}{g(0)}.$$

It is necessary, then, to compute $g'''(0)$.

$$(2.24)\qquad \begin{aligned} g'''(0) &= C_{n+1} \sum{}^* \Big(\sum{}_j u_{\nu_j} a_j\Big)^3\\ &= C_{n+1}(n-3)! \sum u_\nu^3 \Big[n^2 \sum a_j^3 - 3n\Big(\sum a_j\Big) + 2\Big(\sum a_j\Big)^3\Big]. \end{aligned}$$

Note that the rejection rule depends on $\sqrt{b_1}$, the coefficient of skewness through the equations

$$(2.25)\qquad \sum u_\nu^3 = \frac{\sum (y_i - \bar{y})^3}{n^{3/2} s^3} = \frac{\frac{1}{n} \sum (x_i - \bar{x})^3}{\sqrt{n} \left[\frac{1}{n} \sum (x_i - \bar{x})^2\right]^{3/2}} = \frac{\sqrt{b_1}}{\sqrt{n}}.$$

Let $\mu_3\,(a_1, \cdots, a_n)$, to be denoted by $\mu_3\,(a)$, be defined by

(2.26) $$n^3\mu_3(a) = n^2 \sum a_j^3 - 3n(\sum a_j^2)(\sum a_j) + 2(\sum a_j)^3.$$

According to (2.23), (2.24), and (2.25), the rejection criterion (2.22) becomes

(2.27) $$\sqrt{b_1}\, \mu_3(a) > K_1.$$

We formalize the rejection criterion in

THEOREM 2.1. *Under the assumptions of model A, the locally best invariant test of size α for testing H_0 against H_1 is: if $\mu_3(a) > 0$, reject H_0 whenever $\sqrt{b_1} > K_1$; if $\mu_3(a) < 0$, reject H_0 whenever $\sqrt{b_1} < K_1$; where K_1 is chosen to make the test of size α. Suppose $a_j = 0$ for $j = k + 1, \cdots, n$. Then if $n > 2k$, the test which rejects when $\sqrt{b_1} > K_1$ is the locally best invariant test of its size, uniformly in those $(a_1, \cdots, a_k)$ for which $a_j \geqq 0$ for $j = 1, \cdots, k$.*

PROOF. The first assertion follows from the previous discussion. To prove the last assertion, it will be shown that if $a_j = 0$ for $j \geqq k + 1$, if $n > 2k$, and if $a_j \geqq 0$ for all j with at least one $a_j > 0$ from the assumptions of model A), then $\mu_3(a) > 0$. Let $\bar{a}$ denote $(1/k) \sum a_j$. Then

(2.28) $$a_j^3 - 2\bar{a}a_j^2 + \bar{a}^2 a_j = (a_j - \bar{a})^2 a_j \geqq 0.$$

Summing over j, we have

(2.29) $$\sum a_j^3 \geqq 2\bar{a} \sum a_j^2 - k\bar{a}^3.$$

Substituting this inequality into equation (2.26), we find

(2.30) $$\begin{aligned} n^2\mu_3(a) &\geqq (2n^2 - 3nk)\bar{a} \sum a_j^2 - (n^2k - 2k^3)\bar{a}^3 \\ &\geqq k(n - k)(n - 2k)\bar{a}^3 > 0. \end{aligned}$$

The proof is complete.

This theorem can be reworded in the terminology of the problem of the rejection of outliers. If k is allowed to represent the number of spurious observations whose means all may have shifted to the right (model A), the theorem states that the rule which rejects the null hypothesis when the sample coefficient of skewness $\sqrt{b_1}$ is too large, is the locally best invariant test uniformly in the lengths of the shifts of the means, provided that $k < n/2$, that is, provided that there are less than 50% outliers.

We shall now proceed to derive the locally best unbiased invariant test. Using the fact that the first two derivatives with respect to Δ of the density of $Y_2, \cdots, Y_{n-1}$ vanish identically, one may follow the steps in the Neyman-Pearson theory to arrive at the conclusion that the rejection region ω of the locally best unbiased test consists of points such that

(2.31) $$\left.\frac{\partial^4}{\partial\Delta^4} \log f\right|_{\Delta=0} > K_1 \left.\frac{\partial^3}{\partial\Delta^3} \log f\right|_{\Delta=0} + K_2,$$

where f represents the density of $Y_2, \cdots, Y_{n-1}$ and where K_1 and K_2 are constants chosen so that

$$\int_{\omega} f\Big|_{\Delta=0} dx = \alpha, \qquad \int_{\omega} \frac{\partial^3 f}{\partial \Delta^3}\Big|_{\Delta=0} dx = 0. \tag{2.32}$$

In order to derive this test we need first to evaluate the fourth derivative of the logarithm of the density of $Y_2, \cdots, Y_{n-1}$, at $\Delta = 0$. It is easy to see that

$$\frac{\partial^4}{\partial \Delta^4} \log f_{Y_2,\cdots,Y_{n-1}}(y_2, \cdots, y_{n-1}|\Delta)\Big|_{\Delta=0} = \frac{g''''(0)}{g(0)} - 3\left[\frac{g''(0)}{g(0)}\right]^2. \tag{2.33}$$

It is necessary then to compute $g''''(0)$.

$$\begin{aligned} g''''(0) &= C_{n+2} \textstyle\sum^* (\sum u_{\nu_j} a_j)^4 \\ &= C_{n+2}(n-4)! \textstyle\sum u_i^4[(n^3+n^2)(\sum a_j^4) - 4(n^2+n)(\sum a_j^3)(\sum a_j) \\ &\quad - 3(n^2-n)(\sum a_j^2)^2 + 12n(\sum a_j^2)(\sum a_j)^2 - 6(\sum a_j)^4] + c_1, \end{aligned} \tag{2.34}$$

where c_1 is constant, depending on $a_1, \cdots, a_n$, whose exact value it is not necessary to know since it will be absorbed into the constant K_2 of equation (2.31). Note that the rejection rule depends on b_2, the coefficient of kurtosis, through the equation,

$$\sum_{i=1}^{n} u_i^4 = \frac{\frac{1}{n}\sum(x_i - \bar{x})^4}{n\left[\frac{1}{n}\sum(x_i - \bar{x})^2\right]^2} = \frac{1}{n} b_2. \tag{2.35}$$

Let $k_4(a_1, a_2, \cdots, a_n)$, denoted simply by $k_4(a)$, be defined as the fourth k- statistic,

$$\begin{aligned} n(n-1)(n-2)(n-3)\, k_4(a) &= [(n^3+n^2)(\textstyle\sum a_j^4) - 4(n^2+n)(\sum a_j^3)(\sum a_j) \\ &\quad - 3(n^2-2)(\textstyle\sum a_j^2)^2 + 12n(\sum a_j^2)(\sum a_j)^2 - 6(\sum a_j)^4]. \end{aligned} \tag{2.36}$$

Then, using (2.33), (2.34), and (2.35), the rejection criterion (2.31) becomes

$$b_2 k_4(a) > K_1 \sqrt{b_1}\, \mu_3(a) + K_2, \tag{2.37}$$

where K_1 and K_2 are chosen so that conditions (2.32) are satisfied. The only reason $\mu_3(a)$ is not absorbed into the constant K_1 is that it might be zero.

We will show now that when $k_4(a) \neq 0$, and $\mu_3(a) \neq 0$, the second condition of (2.32) is satisfied if and only if $K_1 = 0$. It is easy to check that this second condition can be written as

$$E(\sqrt{b_1}\, I[b_2 k_4 > K_1 \sqrt{b_1}\, \mu_3 + K_2]|\Delta = 0) = 0, \tag{2.38}$$

where $I[A]$ denotes the indicator of the set A. The conditional distribution of $\sqrt{b_1}$ given b_2 is symmetric about $\sqrt{b_1} = 0$ for each value of b_2. In the above equation, we are taking the expected value of $\sqrt{b_1}$ over the set of points on one side of some line in the $(\sqrt{b_1}, b_2)$ plane.

From this it follows that in order for the expected value (2.38) to be zero (yet leaving some mass on each side of the line so that the first condition of (2.32) may be satisfied), it is necessary and sufficient that the line be parallel to the $\sqrt{b_1}$ axis. This occurs if and only if $K_1 = 0$. (If $\mu_3(a)$ happens to be zero, condition (2.38) is automatically satisfied.) The rejection criterion (2.37) has become

$$b_2 k_4(a) > K_2, \tag{2.39}$$

where K_2 is chosen so that the first condition of (2.32) is satisfied. This rejection criterion is thus seen to be the unique locally best unbiased invariant test.

THEOREM 2.2. *Under the assumptions of model A, the locally best unbiased invariant test of size α for testing H_0 against $\bar{H}$ is: if $k_4(a) > 0$, reject H_0 whenever $b_2 > K_2$; if $k_4(a) < 0$, reject H_0 whenever $b_2 < K_2$; where K_2 is chosen so that the probability of rejecting H_0 when true is α. Suppose that $a_j = 0$ for $j = k + 1, \cdots, n$. Then whenever*

$$n > \frac{1}{2}\,[6k - 1 + (12k^2 - 12k + 1)^{1/2}], \tag{2.40}$$

the test which rejects when b_2 is too large is locally best unbiased invariant, uniformly in $(a_1, \cdots, a_k)$.

PROOF. The first statement in the theorem follows from the previous discussion. To prove the last, we will show that whenever inequality (2.40) is satisfied, then $k_4(a) > 0$. The proof will be based in part on the following lemma.

LEMMA. If $\alpha_i > 0$ *for* $i = 1, \cdots, k$, *and if inequality* (2.40) *is satisfied, then*

$$(n^2 + n)\Big(\sum_{i=1}^{k} \alpha_i^3\Big) - 6n\Big(\sum_{i=1}^{k} \alpha_i^2\Big)\Big(\sum_{i=1}^{k} \alpha_i\Big) + 6\Big(\sum_{i=1}^{k} \alpha_i\Big)^3 > 0. \tag{2.41}$$

PROOF OF THE LEMMA. Using inequality (2.29) with a_i replaced by α_i, we see that the left side of (2.41) is greater than or equal to

$$2n(n + 1 - 3k)\,\bar{\alpha}\Big(\sum \alpha_i^2\Big) - k(n^2 + n - 6k^2)\,\bar{\alpha}^3, \tag{2.42}$$

where $\bar{\alpha}$ represents the average of α_i for $i = 1, \cdots, k$. It is clear from inequality (2.40) that $(n + 1 - 3k) > 0$, so that (2.42) is greater than or equal to

$$\begin{aligned} &2n(n + 1 - 3k)\,\bar{\alpha}\,k\,\bar{\alpha}^2 - k(n^2 + n - 6k^2)\,\bar{\alpha}^3 \\ &\qquad = k(n^2 + n - 6nk + 6k^2)\bar{\alpha}^3. \end{aligned} \tag{2.43}$$

That this expression is always positive follows from the fact that (2.40) is exactly the inequality which will insure that the middle factor on the right side of (2.43) is positive. This completes the proof of the lemma.

The first step in the proof of the fact that inequality (2.40) implies that $k_4(a) > 0$, is to show that in the proof we may restrict all the nonzero a_i to be positive. We shall show that, subject to the inequality (2.40), $k_4(a)$ never increases under the operation of replacing all of the a_i by their absolute values. It is sufficient to show that

$$\begin{aligned} &-4(n^2 + n)\Big(\sum a_i\Big)\Big(\sum a_i^3\Big) + 12n\Big(\sum a_i^2\Big)\Big(\sum a_i\Big)^2 - 6\Big(\sum a_i\Big)^4 \\ &\qquad \geqq -4(n^2 + n)\Big(\sum |a_i|\Big)\Big(\sum |a_i|^3\Big) + 12n\Big(\sum a_i^2\Big)\Big(\sum |a_i|\Big)^2 - 6\Big(\sum |a_i|\Big)^4. \end{aligned} \tag{2.44}$$

If we let $\alpha_i = |a_i|$, and if we let $\sum_1$ and $\sum_2$ denote the summation over those indices i for which a_i is positive and negative respectively, inequality (2.44) becomes

(2.45)

$$4(n^2 + n)[(\textstyle\sum_1 \alpha_i + \sum_2 \alpha_i)(\sum_1 \alpha_i^3 + \sum_2 \alpha_i^3) - (\sum_1 \alpha_i - \sum_2 \alpha_i)(\sum_1 \alpha_i^3 - \sum_2 \alpha_i^3)] + 6[(\sum_1 \alpha_i + \sum_2 \alpha_i)^4 - (\sum_1 \alpha_i - \sum_2 \alpha_i)^4] \geqq 12n(\sum_1 \alpha_i^2 + \sum_2 \alpha_i^2)[(\sum_1 \alpha_i + \sum_2 \alpha_i)^2 - (\sum_1 \alpha_i - \sum_2 \alpha_i)^2].$$

This reduces to showing that

$$(2.46)\qquad (\textstyle\sum_1 \alpha_i)[(n^2 + n)(\sum_2 \alpha_i^3) - 6n(\sum_2 \alpha_i^2)(\sum_2 \alpha_i) + 6(\sum_2 \alpha_i)^3] + (\sum_2 \alpha_i)[(n^2 + n)(\sum_1 \alpha_i^3) - 6n(\sum_1 \alpha_i^2)(\sum_1 \alpha_i) + 6(\sum_1 \alpha_i)^3] \geqq 0,$$

which follows immediately from the lemma.

It remains to be shown that whenever the inequality (2.40) is satisfied, then the minimum of $k_4(a)$ over all variation of a such that $a_i > 0$, with $i = 1, \cdots, k$, is positive. It is sufficient to show this under the restriction that $a_1 = 1$, since a multiplication of all the a_i by a constant would then give the result without this restriction. Taking partial derivatives with respect to $a_2, \cdots, a_k$, we find that there can be a minimum of $k_4(a)$ in the interior of the region $a_i > 0$, for $i = 1, \cdots, k$, only at those values of a for which

$$(2.47)\qquad n(n-1)(n-2)\frac{(n-3)}{4}\frac{\partial}{\partial a_j}k_4(a) = (n^3 + n^2)\, a_j^3 - 3(n^2 + n)(\textstyle\sum a_i)\, a_j^2 - 3n[(n-1)\sum a_i^2 - 2(\sum a_i)^2]a_j - [(n^2 + n)\sum a_i^3 - 6n(\sum a_i^2)(\sum a_i) + (\sum a_i)^3] = 0$$

for $j = 2, 3, \cdots, k$. Consider now any fixed set of positive numbers a_i for $i = 2, \cdots, k$, which form a solution of equations (2.47). Each a_i is then a root of that cubic equation obtained by replacing the symbol a_j in equation (2.47) by x. Since there are at most three roots, there are at most three different values for the a_j in any set of solutions. But note that the signs of the coefficients of the cubic are $+, -, -, -$, the first and second being obviously positive and negative, respectively, the last being negative from the lemma, and the third being negative since inequality (2.40) implies that

$$(2.48)\qquad (n-1)\textstyle\sum a_i^2 - 2(\sum a_i)^2 > 2(k\sum a_i^2 - (\sum a_i)^2) \geqq 0.$$

Hence by Descartes' rule of signs there is exactly one positive root, and all the a_j, for $j = 2, \cdots, k$, must be equal to it. Denote that root by b. Substitution of $a_1 = 1$, $a_j = b$, for $j = 2, \cdots, k$, into equation (2.47) yields

$$(2.49)\qquad b^3[n - (k-1)]\,[n^2 + n - 6n(k-1) + 6(k-1)^2] - 3b^2[n^2 + n - 6n(k-1) + 6(k-1)^2] - 3b(n-3)[n - 2(k-1)] - (n-3)(n-2) = 0.$$

The particular positive root b of this equation is less than one, since this function is negative at $b = 0$, and at $b = 1$ it is equal to

$$(2.50)\qquad (n - k)\,[n^2 + n - 6nk + 6k^2] > 0.$$

The term in square brackets is positive from the lemma with $\alpha_i = 1/k$, for $i = 1, \cdots, k$. To prove that the one critical point is a point at which the function is a minimum, we may introduce the following parametrization for an arbitrary line through the point $(1, b, b, \cdots, b)$ contained in the plane $a_1 = 1$,

$$(2.51)\qquad a_j = c_j t + b, \qquad j = 2, \cdots, k,$$

where the c_j are real numbers not all zero and $a_1 = 1$. We substitute this into the formula for $k_4(a)$, take two derivatives with respect to t, and evaluate at $t = 0$. We find that

$$(2.52)\qquad n(n-1)(n-2)(n-3)\left.\frac{\partial^2 k_4(a)}{\partial t^2}\right|_{t=0}$$
$$= 12\sum_{j=2}^{k} c_j^2\, n\{b^2[n-(k-1)][n+1-2(k-1)] - 2b[n+1-2(k-1)] - (n-3)\}$$
$$- 24\Big(\sum_{j=2}^{k} c_j\Big)^2\{b^2[n-(k-1)][2n-3(k-1)] - 2b[2n-3(k-1)] - (n-3)\}.$$

This will be positive irrespective of the c_j, for $j = 2, \cdots, k$, provided they are not all zero, if and only if the coefficient of $\sum c_j^2$ is greater than $(k - 1)$ times the coefficient of $(\sum c_j)^2$. This inequality becomes

$$(2.53)\qquad b^2[n-(k-1)][n^2+n-6n(k-1)+6(k-1)^2]$$
$$- 2b[n^2+n-6n(k-1)+6(k-1)^2] - (n-3)[n-2(k-1)] > 0.$$

We multiply this inequality by b and substitute the value of b^3 found from equation (2.49), to find that we must show that

$$(2.54)\qquad b^2[n^2+n-6n(k-1)+6(k-1)^2]$$
$$+ 2b(n-3)[n-2(k-1)] + (n-3)(n-2) > 0.$$

This is true since each term is positive.

Since the only critical point is a point at which the function takes a minimum value, we need only to evaluate $k_4(a)$ at this point to see if it is positive. The value of $n(n-1)(n-2)(n-3)k_4(a)$ at this point is

$$(2.55)\qquad b^4(k-1)(n-k+1)[n^2+n-6n(k-1)+6(k-1)^2]$$
$$- 4b^3(k-1)[n^2+n-6n(k-1)+6(k-1)^2]$$
$$- 6b^2(k-1)(n-3)[n-2(k-1)]$$
$$- 4b(k-1)(n-2)(n-3) + (n-1)(n-2)(n-3).$$

We want to show that this is positive. First, we subtract $b(k-1)$ times equation (2.49). We must show that

$$(2.56)\qquad -b^3(k-1)[n^2+n-6n(k-1)+6(k-1)^2] \\ -3b^2(k-1)(n-3)[n-2(k-1)] \\ -3b(k-1)(n-2)(n-3)+(n-1)(n-2)(n-3)$$

is positive. It is certainly larger than that value obtained when b is replaced by 1. This turns out to be exactly equation (2.50). This completes the proof of the theorem.

Theorem 2.2, reworded in the language of the rejection of outliers, would be as follows. Let k represent the number of spurious observations whose means may have shifted to the right or left (model A). The rule which rejects the null hypothesis when the sample coefficient of kurtosis b_2 is too large, is the locally best unbiased invariant test, uniformly in the lengths of the shifts provided that inequality (2.40) is satisfied. Replacing the $+1$ under the square root sign in (2.40) by a $+3$, we find that inequality (2.40) is satisfied if

$$(2.57)\qquad n \geqq (3+\sqrt{3})k$$

or if $k/n \leqq 21\%$. Thus, the rule which rejects H_0 when b_2 is too large is locally best unbiased invariant uniformly in the lengths of the shifts, provided there are at most 21% outliers.

2.2 *Model* A, *mean known*. We carry over all the assumptions of model A made in the previous section except that here the mean is assumed to be known. We will take the mean to be zero since we are most interested in the application to factorial designs. In the analysis of a 2^n factorial experiment, one may separate out the stochastically independent estimates of the 2^n-1 main effects and interactions. Under the usual assumptions of the analysis of variance, each of these estimates will have a normal distribution with a common unknown variance and, if the null hypothesis is true, a mean zero. The alternative hypothesis allows the means of a few of the estimates of the effects to be different from zero. This problem is seen to be formally identical with the outlier problem when the mean is known to be zero. In this example we are not interested in one-sided tests. This influences the choice of the first invariance restriction below.

We shall only consider those rejection rules which are invariant under the following three transformations: (1) multiplication of any of the X_i by -1; (2) permutation of the subscripts; (3) multiplication of each X_i by a positive constant.

Any rejection rule invariant under (1) must be a function of $|X_1|, |X_2|, \cdots, |X_n|$. If the rejection rule is invariant under (2), it must be a function only of the order statistics $|X|_{(1)}, |X|_{(2)}, \cdots, |X|_{(n)}$. If in addition the rejection rule is invariant under (3), it must be a function only of ratios, $|X|_{(1)}/|X|_{(n)}, \cdots, |X|_{(n-1)}/|X|_{(n)}$. We now proceed to derive the joint distribution of these ratios. Since the distribution does not depend on σ, we shall take $\sigma=1$ from the start to simplify the algebra.

The density of the absolute value of a normal variable with mean Δ and variance 1 is

$$f(x) = \frac{\sqrt{2}}{\sqrt{\pi}} \exp\left[-\frac{x^2}{2} - \frac{\Delta^2}{2}\right] \cosh \Delta x. \tag{2.58}$$

The joint distribution of the order statistics $|X|_{(1)}, \cdots, |X|_{(n)}$ has the density

$$\begin{aligned} &f_{|X|_{(1)},\cdots,|X|_{(n)}}(x_1, \cdots, x_n) \\ &\quad = \left(\frac{2}{\pi}\right)^{n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n} x_i^2 - \frac{1}{2}\Delta^2 \sum_{j=1} a_j^2\right] \sum{}^* \prod_{j=1}^{n} \cosh \Delta a_j X_{\nu_j}, \end{aligned} \tag{2.59}$$

where $\sum^*$ denotes the summation over all permutations $(\nu_1, \nu_2, \cdots, \nu_n)$ of the first n positive integers.

We shall make the transformation

$$\begin{aligned} Y_i &= \frac{|X|_{(i)}}{|X|_{(n)}} && i = 1, \cdots, n-1, \\ V &= |X|_{(n)}, && \\ |X|_{(i)} &= Y_i V && i = 1, \cdots, n-1, \\ |X|_{(n)} &= V. && \end{aligned} \tag{2.60}$$

The Jacobian of this transformation is V^{n-1}. After making this transformation, we integrate V from 0 to ∞, to find the joint density of $Y_1, \cdots, Y_{n-1}$. In the following formulas, the dummy variable y_n will denote one, s_1^2 will denote $n^{-1} \sum y_i^2$, and u_i will denote $\sqrt{n}\, s_1 y_1$.

$$\begin{aligned} &f_{Y_1,\cdots,Y_{n-1}}(y_1, \cdots, y_{n-1}) \\ &\quad = c \exp\left[-\frac{1}{2}\Delta^2 \sum a_j^2\right] \int_0 \exp\left[-\frac{1}{2} n s_1^2 v^2\right] \sum{}^* \prod_{j=1}^{n} \cosh\,(\Delta\, a_j y_{\nu_j} v) v^{n-1}\, dv \\ &\quad = c \exp\left[-\frac{1}{2}\Delta^2 \sum a_j^2\right] s_1^{-n}\, g(\Delta), \end{aligned} \tag{2.61}$$

where

$$g(\Delta) = \sum{}^* \sum_{(\epsilon_1,\cdots,\epsilon_n)} \int_0^\infty \exp\left[-\frac{1}{2}t^2\right] \cosh\,[\Delta t \sum a_j u_{\nu_j} (-1)^{\epsilon_j}] t^{n-1}\, dt, \tag{2.62}$$

where $0 < y_1 < \cdots < y_{n-1} < 1$, where we have made the change of variable $t = \sqrt{n}\, s_1 v$, and where $\sum_{(\epsilon_1,\cdots,\epsilon_n)}$ denotes the summation $\sum_{\epsilon_1=0}^{1} \sum_{\epsilon_2=0}^{1} \cdots \sum_{\epsilon_n=0}^{1}$.

We shall again apply the Neyman-Pearson method to derive the locally best test of the hypothesis $H_0 : \Delta = 0$ against the alternative $\overline{H} : \Delta \neq 0$. As in the previous section one may justify, using the criterion (2.7), differentiation under the integral sign of formula (2.61). Similarly one may justify placing derivatives of any order under the integral sign of the power function, $\beta_\omega(\Delta)$, as defined previously, whatever be the measurable rejection region ω. The details are as in the previous section and need not be presented here.

We will show that the first three derivatives $\beta_\omega(\Delta)$ vanish at the origin for every invariant rejection region ω. Since the derivatives may be placed under the integral sign, it is sufficient to show that

$$\frac{\partial^i}{\partial \Delta^i} \log f_{Y_1,\cdots,Y_{n-1}}(y_1, \cdots, y_{n-1}|\Delta)\Big|_{\Delta=0} = 0 \tag{2.63}$$

identically in $y_1, \cdots, y_{n-1}$, and for $i = 1,2$, and 3. Since

$$\frac{\partial}{\partial \Delta} \log f_{Y_1,\cdots,Y_{n-1}}(y_1, \cdots, y_{n-1}|\Delta) = -\Delta \sum a_j^2 + \frac{g'(\Delta)}{g(\Delta)}, \tag{2.64}$$

and noting that $g(0) = 2^n n! C_{n-1}$, and $g'(0) = 0$, where C_n is defined by (2.17), (2.63) is true for $i = 1$. The second derivative of the log of the density is

$$\frac{\partial^2}{\partial \Delta^2} \log f_{Y_1,\cdots,Y_{n-1}}(y_1, \cdots, y_{n-1})|\Delta) = -\sum a_j^2 + \frac{g(\Delta)g''(\Delta) - g'(\Delta)^2}{g(\Delta)^2}. \tag{2.65}$$

We need to compute $g''(0)$.

$$\begin{aligned} g''(0) &= \sum{}^* \sum_{(\epsilon_1,\cdots,\epsilon_n)} C_{n+1}[\sum_j a_j u_{\nu_j}(-1)^{\epsilon_j}]^2 \\ &= C_{n+1} 2^n \left(\sum_{j=1}^n a_j^2\right)(n-1)! \\ &= \sum a_j^2 g(0). \end{aligned} \tag{2.66}$$

Thus, (2.63) is true for $i = 2$. One more derivative gives

$$\frac{\partial^3}{\partial \Delta^3} \log f_{Y_1,\cdots,Y_{n-1}}(y_1, \cdots, y_{n-1}|\Delta)\Big|_{\Delta=0} = \frac{g'''(0)}{g(0)}. \tag{2.67}$$

But $g'''(0) \equiv 0$ for the same reason that $g'(0) \equiv 0$. Thus, equation (2.63) has been proved for $i = 1, 2$, and 3.

As in the previous section, if we apply the Neyman-Pearson theory we can deduce that the locally best test for testing hypothesis $H_0 : \Delta = 0$ against the alternative $H : \Delta \neq 0$ (or $H_1 : \Delta > 0$), has a rejection region consisting of those points for which

$$\frac{\partial^4}{\partial \Delta^4} \log f_{Y_1,\cdots,Y_{n-1}}(y_1, \cdots, y_{n-1}|\Delta)\Big|_{\Delta=0} \geqq K, \tag{2.68}$$

where K is chosen so that the probability of rejecting H_0, when true, is a fixed value, α, chosen in advance. One easily finds that

$$\frac{\partial^4}{\partial \Delta^4} \log f_{Y_1,\cdots,Y_{n-1}}(y_1, \cdots, y_{n-1}|\Delta)\Big|_{\Delta=0} = \frac{g''''(0)}{g(0)} - 3\left[\frac{g''(0)}{g(0)}\right]^2. \tag{2.69}$$

It is necessary then to compute $g''''(0)$.

$$\begin{aligned} g''''(0) &= C_{n+3} \sum{}^* \left\{ \sum_{(\epsilon_1,\cdots,\epsilon_n)} \left[\sum_j a_j u_{\nu_j}(-1)^{\epsilon_j}\right]^4 \right\} \\ &= C_{n+3}\, 2^n (n-2)! \sum_{\nu=1} u_\nu^4 \{(n+2) \sum a_j^4 - 3(\sum a_j^2)^2\} + b, \end{aligned} \tag{2.70}$$

where the exact value of the constant b does not matter since it will be absorbed into the constant K of equation (2.68). This rejection rule will depend upon

$$b_2' = \frac{\frac{1}{n}\sum X_i^4}{\left(\frac{1}{n}\sum X_i^2\right)^2} \tag{2.71}$$

through the equations

$$\sum u_\nu^4 = \frac{\sum y_i^4}{n^2 s^4} = \frac{\frac{1}{n}\sum X_i^4}{n\left(\frac{1}{n}\sum X_i^2\right)^2} = \frac{1}{n} b_2'. \tag{2.72}$$

Let $k_4'(a_1, \cdots, a_n)$, to be denoted by $k_4'(a)$, be defined by

$$n(n-1)\, k_4'(a) = (n+2)\sum a_j^4 - 3(\sum a_j^2)^2. \tag{2.73}$$

The rejection criterion (2.68) has become

$$b_2' k_4'(a) > K. \tag{2.74}$$

We arrive at the following theorem.

THEOREM 2.3. *Under the assumptions of model* A *with the mean* $\mu = 0$, *the locally best invariant test of size* α *of the hypothesis* H_0 *against the hypothesis* $\bar{H}$ *(or* H_1*) is: if* $k_4'(a) > 0$ *(respectively* < 0*), reject* H_0 *whenever* $b_2' > K$ *(resp.* $<K$*), where* K *is chosen to make the test of size* α. *Suppose that* $a_j = 0$ *for* $j = k+1, \cdots, n$. *If* $n > 3k - 2$, *then the test which rejects* H_0 *when* $b_2' > K$ *is locally best invariant uniformly in the* a_j *for* $j = 1, \cdots, k$.

PROOF. The first statement follows from the previous discussion. To prove the second statement, we shall show that $n > 3k - 2$ implies that $k_4'(a) > 0$ for all values of a.

$$\begin{aligned} k_4'(a) &= (n+2)\sum a_j^4 - 3(\sum a_j^2)^2 \\ &> 3k \sum a_j^4 - 3(\sum a_j^2)^2 \\ &= 3k \sum_{j=1}^{k} \left(a_j^2 - \frac{1}{k}\sum a_j^2\right)^2 \geqq 0, \end{aligned} \tag{2.75}$$

and the proof is complete.

Using the terminology of the rejection of outliers, this theorem states that, provided there are at most 33% outliers, the rule which rejects when b_2' is too large is locally best invariant uniformly in the lengths of the shifts of the means.

2.3 *Model* B. In this section we derive the locally best invariant tests for the existence of outliers under the assumptions of model B. The precise assumptions of model B are the following: $X_1, X_2, \cdots, X_n$ *are independent random variables having normal distributions with a common mean* μ. *We shall assume that there are known numbers,* $a_1, a_2, \cdots, a_n$ *and unknown parameters* $\sigma > 0$ *and* Δ, *and an unknown permutation* $(\nu_1, \nu_2, \cdots, \nu_n)$ *of the first* n *positive integers such that the* var $X_i = \sigma^2 \exp [\Delta a_{\nu_i}]$, *where* $i = 1, \cdots, n$. *It is also assumed that* $\sum (a_i - \bar{a})^2$ *is positive.*

The null hypothesis that the X_i are identically distributed is $H_0 : \Delta = 0$, while the alternative that some have different variances is that $H_1 : \Delta > 0$. Here we are not really interested in the two-sided alternative $\overline{H} : \Delta \neq 0$. Ordinarily in the outlier problem the outliers will have a larger variance than the nonoutliers, so that the nonzero a_i will be positive. Furthermore, the locally best invariant one-sided test will turn out to be the locally best invariant two-sided test, as we shall see.

The problem above is invariant under the following three transformations: (1) permutation of all the subscripts of the X_i; (2) addition of a constant to all the X_i; (3) multiplication by a positive constant of all the X_i. We shall only concern ourselves with rejection rules which are invariant under these three transformations. Invariance under (1) restricts the rejection rules to functions of the order statistics $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$. Invariance under (2) restricts the rejection rules to functions of the differences, $X_{(2)} - X_{(1)}, \cdots, X_{(n)} - X_{(1)}$, while invariance under (3) restricts further the rejection rules to functions of the ratios, $(X_{(2)} - X_{(1)})/(X_{(n)} - X_{(1)}), \cdots, (X_{(n-1)} - X_{(1)})/(X_{(n)} - X_{(1)})$. We shall derive the joint distribution of these ratios. Since this distribution does not depend on μ and σ, we shall take from the start $\mu = 0$ and $\sigma = 1$ in order to simplify the algebra.

The density of the distribution of the order statistics under model B ($\mu = 0$ and $\sigma = 1$) is,

$$\text{(2.76)} \qquad f_{X_{(1)},\cdots,X_{(n)}}(x_1, \cdots, x_n) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left[-\frac{1}{2}\Delta \sum_j a_j\right] \sum{}^* \exp\left[-\frac{1}{2}\sum_j X_{\nu_j}^2 e^{-\Delta a_j}\right],$$

where $x_1 < x_2 < \cdots < x_n$ and where $\sum^*$ denotes the summation over all permutations $(\nu_1, \cdots, \nu_n)$ of the first n positive integers. To compute the density of the differences, $X_{(2)} - X_{(1)}, \cdots, X_{(n)} - X_{(1)}$, we make the transformation (2.4), whose Jacobian is one, and integrate W from $-\infty$ to ∞.

From this we may find the distribution of the ratios by making the transformation (2.6), whose Jacobian is V^{n-2}, and integrating V from 0 to ∞. We obtain

$$\text{(2.77)} \qquad f_{Y_2,\cdots,Y_{n-1}}(y_2, \cdots, y_{n-1}) = c \exp\left[-\frac{\Delta}{2}\sum a_j\right] \left(\sum e^{-\Delta a_j}\right) s^{-(n-1)/2} g(\Delta),$$

where

$$\text{(2.78)} \qquad g(\Delta) = \sum{}^* \left[\sum u_{\nu_j} e^{-\Delta a_j} - \left(\sum u_{\nu_j} e^{-\Delta a_j}\right)^2 \left(\sum e^{-\Delta a_j}\right)^{-1}\right]^{-(n-1)/2},$$

where $0 = y_1 < y_2 < \cdots < y_{n-1} < y_n = 1$ and where

$$\text{(2.79)} \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^n y_i, \qquad s^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2, \qquad u_j = \frac{y_j - \bar{y}}{\sqrt{n}\, s}.$$

In order to apply the Neyman-Pearson theory we need to be able to differentiate the integral of the density (2.77) over a measurable set in R^{n-2}, with respect to Δ, and to pass this derivative beneath the integral sign. As in section 2.1 it is a straightforward task to show that it is possible to pass derivatives of any order under the integral sign for all values of Δ, and for all measurable sets ω.

We shall now show that the first partial derivative of the logarithm of the density (2.77) with respect to Δ, evaluated at $\Delta = 0$, is zero identically in $0 < y_2 < \cdots < y_{n-1} < 1$, thus implying that the first derivative of the power function vanishes at $\Delta = 0$ whatever be the rejection region ω.

We have

$$(2.80) \qquad \frac{\partial}{\partial\Delta} \log f_{Y_2,\cdots,Y_{n-1}}(y_2, \cdots, y_{n-1}|\Delta) = -\frac{1}{2}\sum a_j + \frac{1}{2}\frac{\sum a_j e^{-\Delta a_j}}{\sum e^{-\Delta a_j}} + \frac{g'(0)}{g(0)}.$$

Since $\sum u_j = 0$ and $\sum u_j^2 = 1$, it may be seen that $g(0) = n!$ and $g'(0) = [(n-1)/2](n-1)! \sum a_j$. Hence, from (2.80),

$$(2.81) \qquad \frac{\partial}{\partial\Delta} \log f_{Y_2,\cdots,Y_{n-1}}(y_2, \cdots, y_{n-1}|\Delta)\Big|_{\Delta=0} = 0.$$

In order to evaluate the second derivative of the logarithm of the density (2.77) at $\Delta = 0$, we first find that

$$(2.82) \qquad g''(0) = \frac{(n+1)!}{4} \sum (a_j - \bar{a})^2 \sum u_j^4 + C(a),$$

where $C(a)$ represents an arbitrary constant which may depend upon the values $a_1, a_2, \cdots, a_n$. Thus, using equation (2.35),

$$(2.83) \qquad \frac{\partial^2}{\partial\Delta^2} f_{Y_2,\cdots,Y_{n-1}}(y_2, \cdots, y_{n-1}|\Delta)\Big|_{\Delta=0} = \frac{(n+1)}{4}\frac{1}{n} \sum (a_j - \bar{a})^2 b_2 + C(a)$$

As in the previous sections the Neyman-Pearson method may be applied to derive the locally best invariant test of size α for testing H_0 against H_1, which here consists of points for which

$$(2.84) \qquad \frac{\partial^2}{\partial\Delta^2} \log f_{Y_2,\cdots,Y_{n-1}}(y_2, \cdots, y_{n-1}|\Delta)\Big|_{\Delta=0} > K,$$

where K is chosen so that the size of the test be α. We have the following theorem.

THEOREM 2.4. *Under the assumptions of model* B, *the locally best invariant test of size α for testing H_0 against H_1 is the following: reject H_0 if $b_2 > K$, where K is chosen so that the size of the test will be α. This test is locally best invariant, uniformly in the a_j.*

This theorem states that no matter how many spurious observations there are, the rule which rejects H_0 when b_2 is too large is locally best invariant under model B assumptions. This highlights the possibility of using the coefficient of kurtosis as a test for homogeneity of variance in a single sample! In practice, however, I doubt if one is ever sure enough of the normality assumption to use this test for that purpose. Still, theorem 2.4 gives an optimum property of the b_2 test when used as it usually is, as a test for normality with unknown mean and variance.

2.4 *Strong local optimality.* In the previous sections it was seen that the locally best tests were optimal uniformly in certain configurations of the shifts, a_i, supposedly known. In this section we shall prove a stronger type of local optimality, in which the a_i will be allowed to be unknown.

We shall put $\Delta = 1$ in the preceding statistical hypothesis, and we shall suppose that for some integer k, $a_i = 0$ for $i > k$. Corresponding to any critical region, ω, there will be a power function, $\beta_\omega(a_1, \cdots, a_k)$, which now is considered as a function of the a_i. We have seen that the locally best tests are unique, so that corresponding to any (unbiased) invariant test, ω', of size α, nontrivially distinct from the locally best test, ω, of size α, and on any line through the origin in k-dimensions, there is an open interval containing the origin throughout which $\beta_\omega(a_1, \cdots, a_k) \geqq \beta_\omega(a_1, \cdots, a_k)$, with equality holding only at the origin. However, it might be that for a fixed test, ω', the length of the interval, being dependent on the line chosen, is not bounded away from zero. That this is impossible is the content of the following corollary. Although this strong local optimality holds for all the locally best tests of the previous section, we shall give the statement and proof only for the locally best unbiased invariant test.

COROLLARY TO THEOREM 2.2. *Let ω_0 be the critical region described in the last statement of theorem 2.2, let k satisfy inequality (2.40), and let ω' be any unbiased invariant critical region of size α, which is distinct from ω_0 in the sense that the Lebesgue measure of their symmetric difference is positive. Then there exists a neighborhood, N, of the origin in k-dimensions throughout which $\beta_{\omega_0}(a_1, \cdots, a_k) > \beta_{\omega'}(a_1, \cdots, a_k)$ except at the origin, where equality holds.*

PROOF. It is easy to check that partial derivatives of all orders of $\beta_\omega(a_1, \cdots, a_k)$ exist for every invariant critical region, ω. If the region is also unbiased, the expansion

$$\beta_\omega(\Delta) = \alpha + \Delta^4 \sum_{hijl} a_h a_i a_j a_l \gamma_\omega^{hijl} + \Delta^5 \delta_\omega(a_1, \cdots, a_k, \Delta) \tag{2.85}$$

is valid, where γ_ω^{hijl} is the fourth partial derivative of $\beta_\omega(a_1, \cdots, a_k)$ with respect to a_h, a_i, a_j, and a_l, evaluated at the origin, and where $\delta_\omega(a_1, \cdots, a_k, \Delta)$ is a continuous function. We are to show that there is a number $\epsilon > 0$ such that the difference

$$\beta_{\omega_0}(\Delta) - \beta_{\omega'}(\Delta) = \Delta^4 \sum a_h a_i a_j a_l (\gamma_{\omega_0}^{hijl} - \gamma_{\omega'}^{hijl}) + \Delta^5 (\delta_{\omega_0} - \delta_{\omega'}) \tag{2.86}$$

is positive, provided $(a_1, \cdots, a_k)$ is on the unit sphere in k-dimensions and $0 < |\Delta| < \epsilon$. However, ω_0 has the property that it is the unique test for which the fourth derivative of $\beta(\Delta)$, evaluated at the origin, is maximized, whatever be the point $(a_1, \cdots, a_k)$. This implies that the coefficient of Δ^4 in formula (2.86) is positive for all points $(a_1, \cdots, a_k)$ on the unit sphere, and hence greater than some number $\eta > 0$, on this sphere. The coefficient of Δ^5 in (2.86), being continuous, is bounded in absolute value by some number $B > 0$, for points $(a_1, \cdots, a_k)$ is on the unit sphere and $|\Delta| < 1$. Hence,

$$\beta_{\omega_0}(\Delta) - \beta_{\omega'}(\Delta) \geqq \Delta^4(\eta - \Delta B). \tag{2.87}$$

The proof is completed by choosing $\epsilon = B/\eta$.

3. Multidecision rejection in model B

It is the purpose of this section to discover the analogue of the rejection rule with the optimum multidecision property as proved by Paulson, Murphy, and Kudo, in the case where alternative hypotheses are of the type specified by model B. We shall treat model B type hypotheses when there is at most one outlier, and we shall consider the multidimensional case as was done by Karlin and Truax for model A alternatives. The model B assumptions which will be invoked in this section are the following.

$\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ *are independent p-dimensional random vectors, each having a normal distribution with a common unknown mean vector* $\boldsymbol{\mu}$. We are interested in finding a multiple decision procedure for choosing among the following $n + 1$ hypotheses:

H_0: $\mathbf{X}_1, \cdots, \mathbf{X}_n$ *have a common covariance matrix*

$$E(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})' = \boldsymbol{\Sigma} \tag{3.1}$$

and, for $k = 1, \cdots, n$,

$$\begin{aligned} H_k: E(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})' &= \boldsymbol{\Sigma} \qquad \text{for } i \neq k \text{ and} \\ E(\mathbf{X}_k - \boldsymbol{\mu})(\mathbf{X}_k - \boldsymbol{\mu})' &= \lambda^2\boldsymbol{\Sigma} \end{aligned} \tag{3.2}$$

where $\boldsymbol{\Sigma}$ *is an unknown* $p \times p$ *nonsingular matrix and where* λ *is a known real number greater than one.*

We shall denote by D_i the decision to act as if hypothesis H_i were true, $i = 0, 1, \cdots, u$. Ordinarily, D_i for $i = 1, \cdots, n$ will be the decision to reject the ith observation as the spurious one.

A decision rule here is a subdivision of the space of possible outcomes of the experiment into $n + 1$ regions, $\omega_0, \omega_1, \cdots, \omega_n$, with the understanding that we shall take decision D_i if the observed outcome of the experiment $(\mathbf{X}_1, \cdots, \mathbf{X}_n)$ falls in ω_i.

We shall restrict attention to those decision rules which are (a) invariant under the addition to each $\mathbf{X}_i$ of a constant vector; (b) invariant under the multiplication of each $\mathbf{X}_i$ by a nonsingular matrix.

For any decision rule satisfying the invariance conditions (a) and (b), the probability of taking decision D_j when hypothesis H_k is true, $P(D_j|H_k)$, is independent of the unknown parameters, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. We shall further restrict attention to those decision rules which satisfy the requirements (c) $P(D_k|H_k)$ is independent of k, for $k = 1, 2, \cdots, n$; (d) $P(D_0|H_0) = 1 - \alpha$, where α is a preassigned number $0 < \alpha < 1$.

Thus although we are confronted in reality with a multiple decision problem, we single out one hypothesis, H_0, for special consideration as in the Neyman-Pearson theory. Out of all decision rules which satisfy (a), (b), (c), and (d), we shall seek that rule which maximizes the probability $P(D_k|H_k)$. This rule turns out to be the rejection criterion based upon the maximum of the lengths of the Studentized deviations from the mean, as was found for model A, by Karlin and Truax. More specifically, let $\overline{\mathbf{X}}$, $\mathbf{S}$, and R_j^2 be the sample mean, the sample covariance matrix, and the jth Studentized square residual, respectively;

$$\begin{aligned} \overline{\mathbf{X}} &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i, \\ \mathbf{S} &= \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i-\overline{\mathbf{X}})(\mathbf{X}_i-\overline{\mathbf{X}})', \\ R_j^2 &= (\mathbf{X}_j-\overline{\mathbf{X}})'\,\mathbf{S}^{-1}(\mathbf{X}_j-\overline{\mathbf{X}}). \end{aligned} \tag{3.3}$$

We shall prove the following theorem.

THEOREM 3.1. *Under model* B *assumptions stated above, the decision rule which, out of all decision rules satisfying* (*a*), (*b*), (*c*), *and* (*d*), *maximizes the probability* $P(D_k|H_k)$, $k \neq 0$, *is: take decision* D_0 *whenever* $\max_j R_j^2 < K$, *where* K *is chosen so that condition* (*d*) *is satisfied; take decision* D_k, *for* $k = 1, \cdots, n$, *whenever* $R_k^2 = \max_j R_j^2 > K$. *This rule is best in the above sense uniformly in* λ.

PROOF. Since the performance of rejection rules satisfying (a) and (b) does not depend on the unknown mean, $\boldsymbol{\mu}$, and unknown covariance matrix, $\boldsymbol{\Sigma}$, we shall take $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}$ in what follows, in order to simplify the algebra. Under hypothesis H_k, we have

$$\begin{aligned} &f_{\mathbf{X}_1,\cdots,\mathbf{X}_n}(\mathbf{x}_1,\cdots,\mathbf{x}_n) \\ &\qquad = \left(\frac{1}{2\pi}\right)^{np/2}\frac{1}{\lambda^p}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\mathbf{x}_j'\mathbf{x}_j+\frac{1}{2}\left(1-\frac{1}{\lambda^2}\right)\mathbf{x}_k'\mathbf{x}_k\right\}. \end{aligned} \tag{3.4}$$

To satisfy requirement (a) we restrict attention to those rules which are functions of $\mathbf{Z}_2 = \mathbf{X}_2 - \mathbf{X}_1$, $\mathbf{Z}_3 = \mathbf{X}_3 - \mathbf{X}_1, \cdots, \mathbf{Z}_n = \mathbf{X}_n - \mathbf{X}_1$. To find the joint distribution of $\mathbf{Z}_2, \cdots, \mathbf{Z}_n$ we make the transformation analogous to (2.4), whose Jacobian is one, and integrate $\mathbf{W}$ over R^p. The dummy variable $\mathbf{z}_1$ in the following formulas represents a p-dimensional vector of zeros.

$$\begin{aligned} f_{\mathbf{Z}_2,\cdots,\mathbf{Z}_n}(\mathbf{z}_2,\cdots,\mathbf{z}_n) &= C\left(\frac{1}{\lambda(n-\gamma)}\right)^p\exp\left\{-\frac{1}{2}\sum(\mathbf{z}_i-\bar{\mathbf{z}})'(\mathbf{z}_i-\bar{\mathbf{z}})\right. \\ &\qquad \left.+\frac{n\gamma}{2(n-\gamma)}(\mathbf{z}_k-\bar{\mathbf{z}})'(\mathbf{z}_k-\bar{\mathbf{z}})\right\}, \end{aligned} \tag{3.5}$$

where $\gamma = 1 - 1/\lambda^2$, and $\bar{\mathbf{z}} = (1/n)\sum_{i=1}^{n}\mathbf{z}_i$. With probability one, the vectors $\mathbf{Z}_{n-p+1}, \cdots, \mathbf{Z}_n$ are linearly independent, so that the $p \times p$ dimensional matrix, $\mathbf{A}$, comprised of the column vectors $(\mathbf{Z}_{n-p+1}, \cdots, \mathbf{Z}_n)$ is nonsingular with probability one. To satisfy requirement (b) we restrict attention to those rules which are functions of $\mathbf{Y}_2 = \mathbf{A}^{-1}\mathbf{Z}_2$, $\mathbf{Y}_3 = \mathbf{A}^{-1}\mathbf{Z}_3, \cdots, \mathbf{Y}_{n-p} = \mathbf{A}^{-1}\mathbf{Z}_{n-p}$. To find the joint distribution of $\mathbf{Y}_2, \cdots, \mathbf{Y}_{n-p}$, we make the transformation

$$\begin{aligned} \mathbf{Y}_j &= \mathbf{A}^{-1}\mathbf{Z}_j, \qquad \text{for } j = 2, \cdots, n-p \\ \mathbf{V}_j &= \mathbf{Z}_{n-p+j}, \qquad \text{for } j = 1, \cdots, p \\ & \qquad \text{where } \mathbf{A} = (\mathbf{Z}_{n-p+j}, \cdots, \mathbf{Z}_n) \\ \mathbf{Z}_j &= \mathbf{A}\mathbf{Y}_j, \qquad \text{for } j = 2, \cdots, n-p \\ \mathbf{Z}_{n-p+j} &= \mathbf{V}_j, \qquad \text{for } j = 1, \cdots, n-p \\ & \qquad \text{where } \mathbf{A} = (\mathbf{V}_1, \cdots, \mathbf{V}_p) \end{aligned} \tag{3.6}$$

whose Jacobian is $||\mathbf{A}||^{n-p-1}$, where by $||\mathbf{A}||$ we mean the absolute value of the determinant of the matrix $\mathbf{A}$. We shall integrate out the variables, $\mathbf{V}_1, \cdots, \mathbf{V}_p$. In the following formulas we shall use the dummy variables $\mathbf{y}_1 = \mathbf{0} = \mathbf{A}^{-1}\mathbf{z}_1$, and $\mathbf{y}_j = \mathbf{e}_{j-n+p} = \mathbf{A}^{-1}\mathbf{z}_j$ for $j = n - p + 1, \cdots, n$, where $\mathbf{e}_j$ is the p-dimensional vector with a one in the jth position and zeros elsewhere.

$$\begin{aligned} &f_{\mathbf{Y}_2,\cdots,\mathbf{Y}_{n-p}}(\mathbf{y}_2, \cdots, \mathbf{y}_{n-p}) \\ &\quad = c\left(\frac{1}{\lambda(n-\gamma)}\right)^p \int \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (\mathbf{y}_i - \bar{\mathbf{y}})' \mathbf{A}'\mathbf{A} (\mathbf{y}_i - \bar{\mathbf{y}}) \right. \\ &\qquad \left. + \frac{n\gamma}{2(n-\gamma)} (\mathbf{y}_k - \bar{\mathbf{y}})' \mathbf{A}'\mathbf{A} (\mathbf{y}_k - \bar{\mathbf{y}})\right\} ||\mathbf{A}||^{n-p-1}\, d\mathbf{A}. \end{aligned} \tag{3.7}$$

Now define $\mathbf{u}_1, \cdots, \mathbf{u}_p$ as those column vectors for which $\mathbf{A}' = (\mathbf{u}_1, \cdots, \mathbf{u}_p)$ so that we may write

$$\sum_{i=1}^{n} (\mathbf{y}_i - \bar{\mathbf{y}})' \mathbf{A}'\mathbf{A} (\mathbf{y}_i - \bar{\mathbf{y}}) = \sum_{j=1}^{p} \mathbf{u}_j \mathbf{S}_0 \mathbf{u}_j, \tag{3.8}$$

where $\mathbf{S}_0$ is the nonsingular matrix

$$\mathbf{S}_0 = \sum_{j=1}^{n} (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'. \tag{3.9}$$

After the change of variables

$$\mathbf{t}_j = \mathbf{S}_0^{1/2}\mathbf{u}_j, \qquad \mathbf{u}_j = \mathbf{S}_0^{-1/2}\mathbf{t}_j, \qquad d\mathbf{u}_j = |\mathbf{S}_0|^{-1/2}\, d\mathbf{t}_j \tag{3.10}$$

for $j = 1, \cdots, p$, equation (3.7) becomes

$$\begin{aligned} &f_{\mathbf{Y}_2,\cdots,\mathbf{Y}_{n-p}}(\mathbf{y}_2, \cdots, \mathbf{y}_{n-p}) \\ &\quad = c\,|\mathbf{S}_0|^{-(n-1)/2}\left(\frac{1}{\lambda(n-\gamma)}\right)^p \int \exp\left\{-\frac{1}{2}\sum_{j=1}^{p} \mathbf{t}_j'\mathbf{t}_j \right. \\ &\qquad \left. + \frac{n\gamma}{2(n-\gamma)} \mathbf{r}_k' \mathbf{B}'\mathbf{B}\mathbf{r}_k\right\} ||\mathbf{B}||^{n-p-1}\, d\mathbf{B}, \end{aligned} \tag{3.11}$$

where $\mathbf{B}' = \mathbf{S}_0^{1/2}\mathbf{A}' = (\mathbf{t}_1, \cdots, \mathbf{t}_p)$ and where $\mathbf{r}_k = \mathbf{S}_0^{-1/2}(\mathbf{y}_k - \bar{\mathbf{y}})$. We note that $\mathbf{r}_k'\mathbf{r}_k$ is the square of the length of the kth Studentized deviation from the mean.

$$R_k^2 = \mathbf{r}_k'\mathbf{r}_k = (\mathbf{y}_k - \bar{\mathbf{y}})'\mathbf{S}_0^{-1}(\mathbf{y}_k - \bar{\mathbf{y}}) = (\mathbf{x}_k - \bar{\mathbf{x}})\mathbf{S}^{-1}(\mathbf{x}_k - \bar{\mathbf{x}}). \tag{3.12}$$

There exists a rotation, $\mathbf{P}$, such that $\mathbf{P}'\mathbf{P} = \mathbf{I}$ and $\mathbf{P}\mathbf{r}_k = R_k\mathbf{e}_1$; then $\mathbf{B}\mathbf{r}_k = R_k\mathbf{B}\mathbf{P}'\mathbf{e}_1$. We shall make another change of variables

$$\mathbf{w}_j = \mathbf{P}\mathbf{t}_j, \qquad \mathbf{t}_j = \mathbf{P}'\mathbf{w}_j, \qquad d\mathbf{t}_j = d\mathbf{w}_j, \tag{3.13}$$

for $i = 1, \cdots, p$, and define $\mathbf{C}' = \mathbf{P}\mathbf{B}' = (\mathbf{w}_1, \cdots, \mathbf{w}_p)$. The density (3.11) becomes

$$\begin{aligned} f_{\mathbf{Y}_2,\cdots,\mathbf{Y}_n}(\mathbf{y}_2, \cdots, \mathbf{y}_n) &= c\left(\frac{1}{\lambda(n-\gamma)}\right)^p |\mathbf{S}_0|^{-(n-1)/2} \int \exp\left\{-\frac{1}{2}\sum_{j=1}^p \mathbf{w}_j'\mathbf{w}_j \right. \\ &\quad \left. + \frac{n\gamma R_k^2}{2(n-\gamma)} \mathbf{e}_1'\mathbf{C}\mathbf{C}'\mathbf{e}_1\right\} ||\mathbf{C}||^{n-p-1}\, d\mathbf{C}. \end{aligned} \tag{3.14}$$

Following the method of Paulson we shall consider the distribution of $\mathbf{Y}_2, \cdots \mathbf{Y}_{n-p}$ only, for which the hypotheses $H_0, H_1, \cdots, H_n$ are simple. We shall choose an a priori distribution for the hypotheses H_k, $k = 0, 1, \cdots, n$, and compute the Bayes solution which maximizes the probability of making a correct decision. This solution will automatically satisfy conditions (a) and (b). If it turns out that this solution also satisfies conditions (c) and (d) then this solution will be the optimum one in the sense described in the theorem. We shall give a priori probabilities p_0 to H_0, and p_1 to H_k for $k = 1, \cdots, n$, where $p_0 + np_1 = 1$. According to the method of Wald, the probability of making a correct decision will be maximized if we take for ω_k, where $k = 0, 1, \cdots, n$, the sets

$$\omega_k = \bigcap_{j=1}^n \{p_k f_k(x) \geqq p_j f_j(x)\}, \tag{3.15}$$

where p_j is the prior probability of density f_j. Under H_k, for $k \neq 0$, the density of $\mathbf{Y}_2, \cdots, \mathbf{Y}_{n-p}$ is given by formula (3.14), and under H_0, it is given by the formula (3.14) with λ replaced by one. Hence

(3.16)

$$\begin{aligned} \omega_0 &= \bigcap_{j=1}^n \left\{p_0 c \geqq p_1 \int \exp\left\{-\frac{1}{2}\sum_{i=1}^p w_i'w_i + \frac{n\gamma R_j^2}{2(n-\gamma)} \mathbf{e}_1'\mathbf{C}\mathbf{C}'\mathbf{e}_1\right\} ||\mathbf{C}||^{n-p-1}\, d\mathbf{C}\right\} \\ &= \bigcap_{j=1}^n \{R_j^2 \leqq K\} = \{\max_j R_j^2 \leqq K\}, \\ \omega_k &= \{R_k^2 > K\} \cap \bigcap_{j=1}^n \{R_j^2 \leqq R_k^2\} \\ &= \{\max_j R_j^2 = R_k^2 > K\}. \end{aligned}$$

Choosing p_0 between zero and one is equivalent to choosing K between zero and infinity, so that by a proper choice of p_0 we may satisfy condition (d). It follows at once from the symmetry of this rule that condition (c) is satisfied. Furthermore since this rule does not depend on the value of λ, it is optimum uniformly in λ, and the proof is complete.

4. Outliers in designed experiments

The problem of rejection of outliers in designed experiments is more complex than for the simple experiments mentioned in the earlier sections of this paper, mainly because it is more difficult to spot spurious observations from the raw data. A large value of an observation may be due to a particular set of values of the unknown parameters, and conversely a spurious observation may appear normal if the shift in the mean cancels the contributions to the mean by the unknown parameters. Furthermore, one bad value among the observations will influence the values of many of the estimates of the parameters, and thus spoil the test completely.

In two recent papers, Anscombe [1] and Daniel [2], a rejection rule based on the maximum Studentized square residual has been suggested on intuitive grounds. Since this is the natural extension of the test based on the maximum Studentized deviation from the mean for which an optimum property was discovered by Murphy [9], Paulson [11], and Kudo [7], it is of interest to investigate how well this optimum property will extend for designed experiments.

4.1. *Distribution of the Studentized square residuals.* Let $\mathbf{X}$ represent an n-dimensional random vector with covariance matrix

$$\text{Cov } \mathbf{X} = \sigma^2 \mathbf{I}_n, \tag{4.1}$$

where $\sigma^2 > 0$ is an unknown parameter and where $\mathbf{I}_n$ is used to denote the $n \times n$ identity matrix. The usual assumptions of the general linear hypothesis are that the mean of $\mathbf{X}$ is $\mathbf{A}\boldsymbol{\xi}$ where $\mathbf{A}$ is an unknown $n \times r$ matrix of full rank $r < n$, and $\boldsymbol{\xi}$ is an r-dimensional vector of unknown parameters. The least squares estimate, $\hat{\boldsymbol{\xi}}$, of $\boldsymbol{\xi}$ is that vector $\boldsymbol{\xi}$ which minimizes the sum of squares

$$S^2 = (\mathbf{X} - \mathbf{A}\boldsymbol{\xi})'(\mathbf{X} - \mathbf{A}\boldsymbol{\xi}). \tag{4.2}$$

The estimate, $\hat{\boldsymbol{\xi}}$, may be obtained by equating $(\partial/\partial\xi)S^2 = -2\mathbf{A}'(\mathbf{X} - \mathbf{A}\boldsymbol{\xi})$ to zero and solving for $\boldsymbol{\xi}$, yielding

$$\hat{\boldsymbol{\xi}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}. \tag{4.3}$$

The vector of residuals, to be denoted by $\mathbf{R}$, is

$$\mathbf{R} = \mathbf{X} - \mathbf{A}\hat{\boldsymbol{\xi}} = \mathbf{B}\mathbf{X}, \tag{4.4}$$

where

$$\mathbf{B} = \mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'. \tag{4.5}$$

It is easy to check directly that $\mathbf{B}$ is a projection (that is, $\mathbf{B}$ is symmetric and $\mathbf{BB} = \mathbf{B}$) and has rank $n - r$. Under the above assumptions $E\mathbf{R} = \mathbf{BA}\boldsymbol{\xi} = 0$, and $\text{Cov } \mathbf{R} = E\mathbf{RR}' = \mathbf{B}E\mathbf{XX}'\mathbf{B}' = \sigma^2\mathbf{B}$. According to the spectral theorem, there exists an n-dimensional matrix $\mathbf{P}$ which is orthogonal (that is, $\mathbf{P}'\mathbf{P} = \mathbf{I}_n$) and for which

$$\mathbf{P'BP} = \begin{pmatrix} \mathbf{I}_{n-r} & 0 \\ 0 & 0 \end{pmatrix} = \mathbf{D}. \tag{4.6}$$

It follows that the residual sum of squares is an unbiased estimate of $(n - r)\sigma^2$, since

$$E \min_{\xi} S^2 = E\mathbf{R'R} = \sigma^2 \text{ trace } \mathbf{B} = \sigma^2(n - r). \tag{4.7}$$

This is the situation which obtains when no spurious values have crept into the experiment. We make allowance for spurious values with the more general assumption that the mean of $\mathbf{X}$ is

$$E\mathbf{X} = \mathbf{A}\xi + \sigma\mathbf{a}, \tag{4.8}$$

where $\mathbf{a}$ is a known n-dimensional vector. We shall derive in this section the distribution of a form of the Studentized square residuals under (4.8) and the assumption that $\mathbf{X}$ has a multivariate normal distribution. The distribution of the Studentized residual itself is singular and will not have a density. We shall derive the distribution of a function of the Studentized residuals, which will have a density in $(n - r - 1)$-dimensions and from which the joint distribution of the Studentized square residuals may be computed if desired. To this end we consider the vector variable,

$$\mathbf{Z} = \mathbf{P'R} = \mathbf{P'BX} = \mathbf{DP'X}, \tag{4.9}$$

where $\mathbf{B}$, $\mathbf{P}$, and $\mathbf{D}$ are as defined previously. The vector $\mathbf{Z}$ will have a multivariate normal distribution with mean

$$E\mathbf{Z} = \mathbf{P'B}(\mathbf{A}\xi + \sigma\mathbf{a}) = \sigma\mathbf{P'Ba} = \sigma\mathbf{DP'a} \tag{4.10}$$

and covariance matrix

$$\text{Cov } \mathbf{Z} = \sigma^2\mathbf{D}. \tag{4.11}$$

Thus the last r components of $\mathbf{Z}$ are degenerate at zero, while the first $n - r$ components are independent normals with a common variance σ^2. The first $n - r$ components of $\mathbf{Z}$ have the density

$$f_{Z_1,\cdots,Z_{n-r}}(z_1, \cdots, z_{n-r}) = c \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{z} - \sigma\mathbf{P'Ba})'(\mathbf{z} - \sigma\mathbf{P'Ba})\right\}, \tag{4.12}$$

where $\mathbf{z}$ represents the n-dimensional vector with transpose $\mathbf{z}' = (z_1, \cdots, z_{n-r}, 0, \cdots, 0)$. To get a form of the density of the Studentized variables, we shall make the transformation

$$\begin{aligned} Y_1 &= Z_1/Z_{n-r} & Z_1 &= WY_1 \\ &\vdots & &\vdots \\ Y_{n-r-1} &= Z_{n-r-1}/Z_{n-r} & Z_{n-r-1} &= WY_{n-r-1} \\ W &= Z_{n-r} & Z_{n-r} &= W \end{aligned} \tag{4.13}$$

The Jacobian of this transformation is $|W|^{n-r-1}$. If we let $\mathbf{y}$ denote the n-dimensional vector with transpose $\mathbf{y}' = (y_1, \cdots, y_{n-r-1}, 1, 0, \cdots, 0)$, we may calculate the density of $Y_1, \cdots, Y_{n-r-1}$ as follows.

$$
\begin{aligned}
(4.14)\qquad & f_{Y_1,\cdots,Y_{n-r-1}}(y_1, \cdots, y_{n-r-1}) \\
&= c\int_{-\infty}^{\infty}\left\{\exp - \frac{1}{2\sigma^2}(w\mathbf{y} - \sigma\mathbf{P}'\mathbf{B}\mathbf{a})'(w\mathbf{y} - \sigma\mathbf{P}'\mathbf{B}\mathbf{a})\right\}|w|^{n-r-1}\,dw \\
&= c\exp\left\{-\frac{1}{2}\mathbf{a}'\mathbf{B}\mathbf{a}\right\}s_0^{-(n-r)}\int_0^{\infty}\exp\left\{-\frac{t^2}{2}\right\}\cosh(t\mathbf{u}'\mathbf{a})t^{n-r-1}\,dt,
\end{aligned}
$$

where we have made the change of variable $t = (w/\sigma)\sqrt{\mathbf{y}'\mathbf{y}}$ and where

$$
(4.15)\qquad \begin{aligned} s_0^2 &= \mathbf{y}'\mathbf{y}, \\ u &= s_0^{-1}\mathbf{B}\mathbf{P}\mathbf{y}. \end{aligned}
$$

Strictly speaking we can reconstruct from $Y_1, \cdots, Y_{n-r-1}$ the values of the Studentized residuals only up to relative signs. We cannot determine the actual signs because of our division by Z_{n-r}, but what we have computed is sufficient to determine the Studentized square residuals used in the theorem of the next section.

4.2. *Invariant admissibility of the rejection rule based on the maximum Studentized square residual.* In the following analysis it will be assumed that there is at most one spurious observation, since, as noted by Murphy [9], the optimum property, which we are going to prove, does not extend conveniently to the case where there are two or more spurious observations. We are concerned here with a multidecision problem. The hypotheses we wish to consider are: H_0, that there are no spurious observations and H_1, that the ith observation alone is spurious, $i = 1, \cdots, n$. Let $\mathbf{e}_i$ denote the n-dimensional column vector with a one in the ith position and zeros in all the remaining positions. Using the notation of the previous section the hypotheses $H_0, H_1, \cdots, H_n$ may be written

$$
(4.16)\qquad \begin{aligned} &H_0: && \mathbf{a} = 0, && \\ &H_i: && \mathbf{a} = a_i\mathbf{e}_i, && i = 1, 2, \cdots, n, \end{aligned}
$$

where the a_i are assumed to be nonzero real numbers known only in absolute value. Allowing the signs of the a_i to be unknown will lead us to two-sided tests. We also make the assumptions that no residual has variance zero, and that no two residuals have correlation equal to $+1$, that is,

$$
(4.17)\qquad \begin{aligned} &b_{ii} \neq 0, && i = 1, \cdots, n, \\ &b_{ij}^2 \neq b_{ii}b_{jj}, && i \neq j, \end{aligned}
$$

where b_{ij} is the (i,j)th element of the matrix $\mathbf{B}$. If b_{ii} were equal to zero for some i, there would be absolutely no way of telling whether or not the ith observation is spurious without the use of supplementary information. Especially embarrassing is the situation in which a few of the residuals have correlations very close or equal to plus or minus one. A spurious value which occurs at one of these observations effects very strongly those residuals which have a very strong positive or negative correlation with its residual, making it very difficult to judge which of the observations is the maverick. For a discussion on this point, and an example, see Anscombe [1].

The problem we are considering is invariant under the two operations: (1) addition to the vector **X** of a vector $\mathbf{A}\boldsymbol{\alpha}$ where $\boldsymbol{\alpha}$ is an arbitrary r-dimensional vector; (2) multiplication of the vector **X** by an arbitrary nonzero scalar. As a consequence, we shall only consider rules which are invariant under these two operations. Invariance under the first operation requires that we consider only those rules which are functions of the vector of residuals, **R**. In addition, invariance under the second operation requires that we consider only those rules which are functions of that form of the Studentized residuals found in the previous section, $Y_1, \cdots, Y_{n-r-1}$. Each of the hypotheses, H_1, when applied to the distributions of such invariant rules, is simple, since the distributions are independent of the actual values of $\boldsymbol{\xi}$, σ^2, and the signs of the a_i.

As before we shall let D_i represent the decision to reject the ith observation as spurious, and D_0 to represent the decision to accept all the observations as valid. The analogue of the condition found in Paulson that $P(D_i|H_i)$ be independent of i for $i = 1, \cdots, n$, is no longer a reasonable condition to impose on the decision rule, because the problem is no longer symmetric in the observations, that is, the residuals do not necessarily have equal variances or correlations.

The optimal property of the rule based on the maximum Studentized deviation from the mean as found by Paulson for model A, and in section three for model B, is nothing more nor less than the facts that the rule is (a) symmetric in the observations, and (b) admissible among all invariant rules, being Bayes with respect to an a priori distribution giving positive mass to each of the $n + 1$ simple hypotheses. Thus, the optimality is admissibility among invariant rules. In what follows we shall say that a decision rule is *invariant admissible* for the problem we are considering, if it is invariant, and if $P(D_i|H_i)$ for fixed i cannot be increased by using another invariant decision rule without decreasing $P(D_j|H_j)$ for some $j \neq i$. Since we are no longer interested in symmetry, the analogue of the Paulson optimum property for the problem under consideration will be simply that the rejection rule based on the maximum Studentized residual is invariant admissible. This is the content of the following theorem. Let us denote the residual sum of squares by S_0^2 and jth Studentized square residual by V_j^2, so that

$$\begin{aligned} S_0^2 &= \mathbf{R}'\mathbf{R}, \\ V_j^2 &= R_j^2(b_{jj}S_0^2)^{-1}(n - r), \qquad j = 1, 2, \cdots, n. \end{aligned} \tag{4.18}$$

THEOREM 4.1. *The decision rule which states, take decision D_0 when $\max_j V_j \leq K$, and for $i = 1, \cdots, n$ take decision D_i when $V_i = \max_j V_j > K$, is invariant admissible for the outlier problem when $|a_i| = ab_{ii}^{-1/2}$ where a is a positive real number, the rule being Bayes with respect to an a priori distribution giving equal weights to hypotheses $H_1, \cdots, H_n$. This optimum property holds uniformly in a.*

REMARKS. This theorem can be considered a rather satisfactory extension of the Paulson optimum property in the case where all the residuals have equal variances, that is, all b_{ii} are equal. It is this type of design to which Anscombe restricted his attention; that it still covers a large class of designs is evident in the list of such designs found in Anscombe's paper. For such designs, the above theorem tells us that the rejection rule based on the maximum Studentized square residual is a Bayes solution with respect to the a priori distribution which assigns equal weights to hypotheses specifying *equal* shifts in the mean for the spurious observation.

This may be contrasted with the general statement that the rule based on the maximum Studentized square residual is a Bayes solution with respect to the a priori distribution which assigns equal weights to hypotheses specifying a shift of length $\pm a(\sigma/\sqrt{b_{kk}})$ in an observation whose residual has variance $\sigma\, b_{kk}$. Essentially, we are guarding equally against shifts in the mean, of length inversely proportional to the standard deviations of the residuals, so that a small shift is hypothesized for observations whose residuals have large variances. If the null hypothesis is true, and we do decide to reject, the various alternative hypotheses will be accepted with approximately equal probabilities (depending on the approximate equality of the correlations of the residuals). On the other hand, if hypothesis H_k is true, $k \neq 0$, the vector of residuals gets shifted by an amount $\pm a(\sigma/\sqrt{b_{kk}})\mathbf{B}\mathbf{e}_k = \pm a(\sigma/\sqrt{b_{kk}})\mathbf{b}_k$, where $\mathbf{b}_k$ is the kth column of $\mathbf{B}$, and represents the covariance of $\mathbf{R}$ and its kth component, R_k. Consequently, the Studentized jth residual, $V_j = R_j/s\sqrt{b_{jj}}$, gets shifted by an amount $\pm a\rho_{jk}(\sigma/s)$, where ρ_{jk} represents the correlation between the jth and the kth residuals.

One can deduce from this that when b_{kk} is small relative to the other b_{jj}, a relatively large shift, of length $\pm a(\sigma/\sqrt{b_{kk}})$, in the kth observation will be required in order to reject H_0, and conversely, when b_{kk} is relatively large, a relatively small shift in the kth observation will be sufficient to reject H_0. Furthermore, it is easy to conjecture from the above that $P(D_j|H_k)$ takes on its maximum value for fixed k when $j = k$. This statement is the analogue of one of the unbiasedness properties proved by Kapur [5] for the maximum Studentized deviation from the mean. Its validity in the present situation is still an open problem.

Another way of viewing this situation is as follows. When a residual has a relatively small variance, most of the information provided by the corresponding observation goes into the estimation of the unknown $\boldsymbol{\xi}$, while relatively little is used to estimate the unknown σ^2. It is only by that part of the observation used to estimate σ^2 that we can hope to judge whether or not the corresponding observation is spurious. Thus we should require a larger shift for an observation whose residual has a smaller variance, in order to be able to reject it. This turns out to be the case with the rejection rule of the above theorem; in fact the length of shift required is roughly proportional to the inverse of the standard deviation of the residual. Although these considerations point out the fact that this rejection rule is not easily justifiable in the case that the b_{jj} are not all equal, still it cannot be called an unreasonable rule.

PROOF. First, we shall restrict attention to the invariant decision rules, so that as far as we are concerned the hypotheses H_i are simple. We will show that the above decision rule is a Bayes rule with respect to an a priori distribution giving positive mass to each of the $n+1$ hypotheses H_i where $i = 0, 1, \cdots, n$.

Let us denote the a priori probability of H_i by p_i, $i = 0, 1, \cdots, n$, where $p_i > 0$ for all i and $\sum_{i=0}^{n} p_i = 1$. Wald's method may be applied to find that decision rule which maximizes the probability of making a correct decision, $\sum_{i=0}^{n} p_i P(D_i|H_i)$. Let us denote by $f^{(i)}_{Y_1,\cdots,Y_{n-r-1}}(y_1, \cdots, y_{n-r-1})$ the density of $Y_1, \cdots, Y_{n-r-1}$ under hypothesis H_i, for $i = 0, 1, \cdots, n$. Then the Bayes decision rule tells us to take decision D_i, whenever the sample point falls in ω_i, where for $i = 0, 1, \cdots, n$,

$$\omega_i = \bigcap_{\substack{j=0 \\ j\neq i}}^{n} \left\{p_i f^{(i)}_{Y_1,\cdots,Y_{n-r-1}}(y_1, \cdots, y_{n-r-1}) > p_j f^{(j)}_{Y_1,\cdots,Y_{n-r-1}}(y_1, \cdots, y_{n-r-1})\right\}. \tag{4.19}$$

The density of $Y_1, \cdots, Y_{n-r-1}$ may be found from equation (4.14) for each hypothesis. The function,

$$g(x) = \int_0^\infty \exp\left\{-\frac{t^2}{2}\right\} \cosh(tx) t^{n-r-1}\, dt, \tag{4.20}$$

is an even function of x, increasing for $0 \leqq x < \infty$.
We may write

$$f^{(0)}_{Y_1,\cdots,Y_{n-r-1}}(y_1, \cdots, y_{n-r-1}) = c s_0^{-(n-r)} g(0), \tag{4.21}$$

and, for $j = 1, 2, \cdots, n$,

$$f^{(j)}_{Y_1,\cdots,Y_{n-r-1}}(y_1, \cdots, y_{n-r-1}) = c s_0^{-(n-r)} \exp\left\{-\frac{a_j^2}{2} \mathbf{e}_j'\mathbf{B}\mathbf{e}_j\right\} g(a_j\mathbf{u}'\mathbf{e}_j). \tag{4.22}$$

Noting that $\mathbf{e}_j'\mathbf{B}\mathbf{e}_j = b_{jj}$, and denoting $\mathbf{u}'\mathbf{e}_j$ by u_j, we may write formulas (4.19) as

$$\omega_0 = \bigcap_{j=1}^{n} \left\{p_0 g(0) > p_j \exp\left(-\frac{1}{2} a_j^2 b_{jj}\right) g(a_j u_j)\right\}, \tag{4.23}$$

and, for $i = 1, \cdots, n$,

$$\omega_i = \left\{p_0 g(0) < p_i \exp\left(-\frac{1}{2} a_i^2 b_{ii}\right) g(a_i u_i)\right\} \tag{4.24}$$

$$\bigcap_{\substack{j=1 \\ j\neq i}}^{n} \left\{p_i \exp\left(-\frac{1}{2} a_i^2 b_{ii}\right) g(a_i u_i) > p_j \exp\left(-\frac{1}{2} a_j^2 b_{jj}\right) g(a_j u_j)\right\}.$$

All these rules are invariant admissible in the different problems for which they are designed, but most of them will lead to difficult computations in practical situations. However, a great simplification occurs when $p_i \exp\{-(1/2)a_i^2 b_{ii}\}$ is independent of i, for $i = 1, 2, \cdots, n$. Let us choose, therefore,

$$p_i = c_0 \exp\left\{\frac{1}{2} a_i^2 b_{ii}\right\}, \tag{4.25}$$

for $i = 1, 2, \cdots, n$ where $c_0 > 0$ is chosen so that $p_0 = 1 - \sum_{i=1}^{n} p_i$ is positive. Then

$$\omega_0 = \{\max_j (a_j u_j) < K\}, \tag{4.26}$$

and, for $i = 1, \cdots, n$,

$$\omega_i = \{(a_i u_i)^2 = \max_j (a_j u_j)^2 > K\}, \tag{4.27}$$

where K is the positive root of the equation $g(K^{1/2}) = p_0 g(0) c_0^{-1}$. K will exist if we further restrict $c_0 < p_0$, or $0 < c_0 < (1 + \sum \exp \{(1/2) a_j^2 b_{jj}\})^{-1}$. A choice of c_0 in this range is equivalent to a choice of K on the positive axis. The above formulas define a class of decision rules, one for each choice of nonzero $a_1, \cdots, a_n$. The particular decision rule mentioned in the theorem will emerge when $a_i^2 = a^2 b_{ii}^{-1}$, since then

$$\frac{a_i^2 \mathbf{u}_i^2}{a^2} = \frac{(\mathbf{u}'\mathbf{e}_i)^2}{b_{ii}} = \frac{(\mathbf{z}'\mathbf{P}'\mathbf{B}\mathbf{e}_i)}{\mathbf{z}'\mathbf{z} b_{ii}} = \frac{1}{n - r} V_i^2. \tag{4.28}$$

We note from equation (4.25) that this particular decision rule is Bayes with respect to an a priori distribution giving equal weights to the hypotheses H_i, $i = 1, 2, \cdots, n$. This completes the proof of the theorem.

5. Results of a sampling experiment

In section two the suggestion was made of the use of the coefficient of skewness and the coefficient of kurtosis for the rejection of outliers. These rules were seen to have a theoretical optimum property and no consideration was given to practical applicability. There are several points which should be mentioned.

First, there is a lack of tables of the percentage points for both the coefficients of skewness and the coefficient of kurtosis. For large n ($n > 25$ for $\sqrt{b_1}$, and $n > 200$ for b_2), tables may be found in Biometrika Tables for Statisticians [12]. But as far as the author is aware, no accurate tables are in existence for smaller values of n. As a by-product of the sampling experiment described in more detail below, certain percentage points for the distributions of $\sqrt{b_1}$ and b_2 have been approximated and may be found in tables I and II. It is hoped that these tables will prove useful until more accurate tables become available.

TABLE I

ESTIMATED CUTOFF POINTS FOR THE DISTRIBUTION OF $\sqrt{b_1}$

Based on $2N_n$ samples. (Advantage was taken of the symmetry of the distribution.) The entries for $n = 25$ are the most inaccurate, being based on a sample of size 2000. The 95% confidence intervals for the 1%, 5%, and 10% cutoff points for $n = 25$ were respectively (.94, 1.08), (.66, .72), and (.51, .57). The Biometrika tables [12] for the 1% and 5% cutoff points, for $n = 25$, give respectively 1.061 and .711, which fall within the confidence intervals.

α \ n	5	10	15	20	25
.01	1.34	1.31	1.20	1.11	.98
.05	1.05	.92	.84	.79	.68
.10	.82	.71	.66	.59	.54

TABLE II

ESTIMATED CUTOFF POINTS FOR THE DISTRIBUTION OF b_2

Based on N_n samples. The entries for $n = 25$ are the most inaccurate being based on a sample of size 1000. The 95% confidence intervals for the 1%, 5%, and 10% cutoff points for $n = 25$, were respectively (4.60, 5.60), (3.81, 4.19), (3.48, 3.72).

α \ n	5	10	15	20	25
.01	3.11	4.83	5.08	5.23	5.00
.05	2.89	3.85	4.07	4.15	4.00
.10	2.70	3.40	3.54	3.65	3.57

Second, the optimum property proved for the $\sqrt{b_1}$ and b_2 tests is not a very convincing one. If, for example, we compare $\sqrt{b_1}$ with the Studentized maximum deviation from the mean suggested by Grubbs [4], and the statistic R_{10} suggested by Dixon [3], on the basis of the power of the tests under model A assumptions when there is at most one outlier, we are certain that for sufficiently small shifts in the mean of the spurious observation the rule based on $\sqrt{b_1}$ is best. However, for small shifts in the mean none of these tests will be any good, and for sufficiently large shifts all the tests will have power close to one. It is more important to investigate the relative behavior of these tests for "medium-sized" shifts. This has been done by actually sampling from the normal distribution to obtain estimates of the behavior.

For this purpose, a tape of 25,000 random normal deviates, which had been checked with special reference to the behavior at the tails of the distribution was provided at Bell Telephone Laboratories, Murray Hill, New Jersey, for use on the IBM 704.

Successive samples of size n were taken from a normal population with mean zero and variance one, and to a fixed member of each sample was added, successively, the nonnegative integers $\lambda = 0, 1, 2, \cdots, 15$. The values of n chosen for this study were $n = 5, 10, 15, 20$, and 25. For each fixed value of n there is an upper limit to the number of samples that can be obtained if each of the 25,000 random normal deviates is used at most once for each distribution. If we let N_n denote the number of samples of size n that we have taken, then $N_5 = 5{,}000$, $N_{10} = 2{,}500$, $N_{15} = 1{,}650$, $N_{20} = 1{,}250$, $N_{25} = 1{,}000$. For each fixed value of n, and each fixed λ, the following statistics have been computed for each sample:

One-sided

$$\begin{aligned} \sqrt{b_1} &= \sum (X_i - \overline{X})^3/ns^3 \\ SMD &= (X_{(n)} - \overline{X})/s \\ R_{10} &= (X_{(n)} - X_{(n-1)})/(X_{(n)} - X_{(1)}) \end{aligned}$$

Two-sided

$$b_2 = \sum (X_i - \overline{X})^4/ns^4$$
$$SMD^{(2)} = \max [SMD, (\overline{X} - X_{(1)})/s]$$
$$R_{10}^{(2)} = \max [R_{10}, (X_{(2)} - X_{(1)})/(X_{(n)} - X_{(1)})]$$

where $\overline{X}$ and s^2 represent the sample mean and standard deviation of the sample, $X_1, \cdots, X_n$, and where $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ represent the order statistics.

This study somewhat overlaps the work done by Dixon [3]; however, the number of samples per distribution is increased here over that of Dixon by a factor of about 10, and so the results may be considered more accurate. On the other hand, Dixon's point of view is slightly different from that presented here. He has tabled and graphed the probability of rejecting the spurious observation. The result of Paulson says that the statistic *SMD* is the best invariant with respect to this criterion for all values of the shift λ. Unfortunately, Dixon's sample sizes were too small for him to be able to detect this fact experimentally.

In this paper we are considering rejection of outliers from the point of view of hypothesis testing. The power function of the test of the null hypothesis that no observation is spurious has been computed and tabled as a function of λ. It was found that for small values of n there is no difference to approximately two decimal places between the power functions of the one-sided tests or between the power functions of the two-sided tests. For $n = 5$ these power functions may be found in tables III and IV.

TABLE III

POWER OF THE ONE-SIDED TESTS, $\sqrt{b_1}$, SMD, R_{10}, $n = 5$

Based on 5000 samples.

α \ λ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
.01	.01	.03	.06	.12	.20	.29	.39	.49	.59	.69	.77	.83	.88	.92	.95
.05	.06	.12	.24	.40	.57	.72	.83	.91	.96	.98	.99	1.00			
.10	.12	.22	.39	.60	.77	.88	.95	.99	1.00						

TABLE IV

POWER OF THE TWO-SIDED TESTS, b_2, $SMD^{(2)}$, $R_{10}^{(2)}$, $n = 5$

Based on 5000 samples.

α \ λ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
.01	.01	.02	.03	.06	.11	.17	.24	.32	.40	.48	.56	.64	.71	.77	.82
.05	.05	.07	.13	.24	.37	.50	.63	.74	.83	.90	.94	.97	.98	.99	1.00
.10	.10	.14	.25	.40	.57	.72	.83	.91	.96	.98	.99	1.00			

However, for larger values of n, it was found that there are significant differences between the power functions of the various tests. The difference seems to increase slowly with n. For $n = 25$, these power functions may be found in tables V and VI. It is seen in table V that for the one-sided tests SMD is significantly

TABLE V

POWER OF THE ONE-SIDED TESTS, $n = 25$

Based on 1000 samples.

1% level of significance

λ	1	2	3	4	5	6	7	8
$\sqrt{b_1}$	.01	.07	.23	.52	.80	.94	.99	1.00
SMD	.01	.06	.22	.53	.82	.96	.99	1.00
R_{10}	.01	.05	.18	.45	.75	.92	.98	.99

5% level of significance

λ	1	2	3	4	5	6	7	8
$\sqrt{b_1}$	.06	.17	.38	.69	.89	.98	1.00	
SMD	.06	.16	.42	.75	.94	.99	1.00	
R_{10}	.06	.13	.36	.67	.90	.98	.99	1.00

10% level of significance

λ	1	2	3	4	5	6	7	8
$\sqrt{b_1}$	.11	.23	.49	.76	.93	.99	1.00	
SMD	.10	.24	.55	.83	.97	.99	1.00	
R_{10}	.12	.20	.47	.78	.94	.99	1.00	

better than $\sqrt{b_1}$ or R_{10} at the 5% and 10% significance levels, and that there is not much difference between $\sqrt{b_1}$ and R_{10}. At the 1% level $\sqrt{b_1}$ and SMD are approximately the same and both are better than R_{10}. In table VI it is seen that for the two-sided tests b_2 and $SMD^{(2)}$ are approximately equivalent at the 1%, 5% and 10% significance levels, $SMD^{(2)}$ being slightly better, and that both are significantly better than $R_{10}^{(2)}$. However, it should be pointed out that for $n = 25$ the power of R_{10} can be increased slightly by modifying it according to the suggestions of Dixon, since for large values of n the range is not as efficient as, for example, $X_{(n-1)} - X_{(2)}$ for use in estimating σ.

In table VI the cutoff points for all three statistics had to be estimated by a sampling procedure, thus yielding an extra source of error of the power tables. The exact cutoff points for the $SMD^{(2)}$ test and certain generalizations of it have been computed by Dr. C. Quesenberry at the University of North Carolina, as part of his Ph.D. thesis.

TABLE VI

POWER OF THE TWO-SIDED TESTS, $n = 25$

Based on 1000 samples.

1% level of significance

λ	1	2	3	4	5	6	7	8
b_2	.02	.04	.19	.48	.79	.95	.99	1.00
$SMD^{(2)}$	.01	.04	.18	.47	.78	.94	.99	1.00
$R_{10}^{(2)}$	.01	.04	.14	.36	.66	.88	.97	.99

5% level of significance

λ	1	2	3	4	5	6	7	8
b_2	.05	.12	.36	.68	.89	.98	1.00	
$SMD^{(2)}$	.05	.13	.35	.69	.91	.99	1.00	
$R_{10}^{(2)}$	.05	.11	.29	.60	.84	.96	.99	1.00

10% level of significance

λ	1	2	3	4	5	6	7	8
b_2	.10	.20	.46	.76	.94	.99	1.00	
$SMD^{(2)}$	.11	.20	.47	.78	.95	.99	1.00	
$R_{10}^{(2)}$	.10	.17	.39	.69	.90	.98	.99	1.00

Because the test based on b_2 has an optimum property which is uniform in the number of spurious observations up to 21% of the total, it should be especially useful for repeated rejections. For this, the b_2 test is applied to the data; if the null hypothesis is rejected, the observation farthest from the mean is rejected from the data, and the b_2 test is applied again; and so on. Sequential rejection plans have been suggested by Grubbs [4] and Murphy [9], using modified forms of SMD, and by Dixon [3], using modified forms of R_{10}. The advantage of the sequential plan based on $\sqrt{b_1}$ or b_2, is that no modified forms are necessary and that, consequently, one set of tables is sufficient for all the rejections.

The use of $SMD^{(2)}$ for repeated rejections is not recommended because of what Murphy refers to as the masking effect: for small sample sizes and a small level of significance ($n \leqq 15$ and $\alpha < .05$, for example) and for two spurious observations equally far from the mean, neither will be rejected. It may easily be guessed that b_2 suffers from a slight masking effect also. Thus it is of interest to determine whether b_2 is significantly better than $SMD^{(2)}$ in this respect. A second sampling experiment was performed, this time with two spurious observations, in the situation where the masking effect should be the greatest, namely when each spurious observation is shifted an equal distance in the positive direction.

The effect on the probability of rejecting the null hypothesis may be found in tables VII and VIII. For $n = 5$ and $n = 10$, the masking effect is very strong,

TABLE VII

POWER OF b_2 AND $SMD^{(2)}$ WHEN THERE ARE TWO SPURIOUS OBSERVATIONS BOTH SHIFTED BY A LENGTH $+\lambda\sigma$, $n = 15$

Based on 1650 samples.

1% level of significance

λ	1	2	3	4	5	6	7	8	9	10	11	12
b_2	.02	.03	.05	.09	.12	.16	.23	.33	.46	.61	.74	.84
$SMD^{(2)}$	.02	.03	.06	.08	.09	.09	.08	.07	.06	.05	.04	.03

5% level of significance

λ	1	2	3	4	5	6	7	8	9	10	11	12
b_2	.06	.09	.17	.32	.54	.75	.91	.98	1.00			
$SMD^{(2)}$	.06	.09	.18	.26	.33	.38	.42	.44	.46	.47	.47	.48

10% level of significance

λ	1	2	3	4	5	6	7	8	9	10	11	12
b_2	.11	.16	.32	.56	.80	.95	.99	1.00				
$SMD^{(2)}$	.11	.16	.29	.41	.53	.62	.71	.78	.84	.89	.93	.96

TABLE VIII

POWER OF b_2 AND $SMD^{(2)}$ WHEN THERE ARE TWO SPURIOUS OBSERVATIONS BOTH SHIFTED BY A LENGTH $+\lambda\sigma$, $n = 25$

Based on 1000 samples.

1% level of significance

λ	1	2	3	4	5	6	7	8	9	10
b_2	.02	.04	.16	.41	.73	.93	1.00			
$SMD^{(2)}$	.02	.04	.12	.25	.42	.56	.69	.84	.93	.97

5% level of significance

λ	1	2	3	4	5	6	7	8
b_2	.06	.14	.37	.71	.95	1.00		
$SMD^{(2)}$	.06	.13	.34	.58	.81	.96	.99	1.00

10% level of significance

λ	1	2	3	4	5	6	7
b_2	.11	.22	.50	.84	.97	1.00	
$SMD^{(2)}$	.11	.22	.46	.75	.93	.99	1.00

both for b_2 and $SMD^{(2)}$; in fact, for both tests the probability of rejection tends to zero as λ tends to infinity (we shall say that the tests are inconsistent) at levels 1%, 5%, and 10%. For $n = 15$, $SMD^{(2)}$ is inconsistent at levels 1% and 5%; and even though consistent at the 10% level, it has a poor behavior in comparison to b_2. At $n = 25$, both tests are consistent at levels 1%, 5%, and 10% but the b_2 test is significantly better. In addition, by comparing table VIII with table VI the power of the test based on b_2 is found to have increased as the number of spurious observations has increased.

The author is indebted to the Bell Telephone Laboratories at Murray Hill, New Jersey, for the very generous permission to use their IBM 704 for the computation of the tables in the last section. He would like especially to express his gratitude to Miss Marilyn Huyett of Bell Telephone Laboratories for her expert help in programming the problem. Thanks are also due to Drs. J. W. Tukey, H. A. David, Allan Birnbaum, and Frank Anscombe for many helpful discussions.

REFERENCES

[1] F. J. ANSCOMBE, "Rejection of outliers," *Technometrics*, Vol. 2 (1960), pp. 123–147.
[2] C. DANIEL, "Locating outliers in factorial experiments," *Technometrics*, Vol. 2 (1960), pp. 149–156.
[3] W. J. DIXON, "Analysis of extreme values," *Ann. Math. Statist.*, Vol. 21 (1950), pp. 488–506.
[4] F. E. GRUBBS, "Sample criteria for testing outlying observations," *Ann. Math. Statist.*, Vol. 21 (1950), pp. 27–58.
[5] M. N. DAPUR, "A property of the optimum solution suggested by Paulson for the k-sample slippage problem for the moral distribution," *J. Indian Soc. Agric. Statist.*, Vol. 9 (1957), pp. 179–190.
[6] S. KARLIN and D. TRUAX, "Slippage problems," *Ann. Math. Statist.*, Vol. 31 (1960), pp. 296–324.
[7] A. KUDO, "On the testing of outlying observations," *Sankhyā*, Vol. 17 (1956), pp. 67–76.
[8] M. LOÈVE, *Probability Theory*, New York, Van Nostrand, 1955.
[9] R. B. MURPHY, "On tests for outlying observations," unpublished Ph.D. thesis, Princeton University, 1951.
[10] J. NEYMAN and E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses," *Statist. Res. Mem.*, Vol. 1 (1938), pp. 1–37.
[11] E. PAULSON, "An optimum solution to the k-sample slippage problem for the normal distribution," *Ann. Math. Statist.*, Vol. 23 (1952), pp. 610–616.
[12] E. S. PEARSON and H. O. HARTLEY (eds.), *Biometrika Tables for Statisticians*, Vol. I, Cambridge, Cambridge University Press, 1954 (4th ed.).