TECHNOMETRICS VOL. 11, NO. 3 AUGUST 1969

Investigation of Rules for Dealing With Outliers in Small Samples from the Normal Distribution: I: Estimation of the Mean

IRWIN GUTTMAN AND DENNIS E. SMITH*

University of Wisconsin

The performance of three rules for dealing with outliers in small samples of size n from the normal distribution $N(\mu, \sigma^2)$ are investigated when the primary objective of sampling is to obtain an accurate estimate of μ. It is assumed that at most one observation in the sample may be biased, arising from either $N(\mu + a\sigma, \sigma^2)$ or $N(\mu, (1 + b)\sigma^2)$. Performance of each rule is measured in terms of "Protection", the fractional decrease in the Mean Square Error (MSE) obtained by using the rule when a biased observation actually is present in the sample. Although numerical results have been obtained for $n \leq 10$ when σ^2 is known, computational difficulties have prevented evaluation of protections when σ^2 is unknown except when $n = 3$.

1. INTRODUCTION

The problem of how to deal with data which contain "outliers", i.e., observations which look suspicious in some way, has long been a source of concern to experimenters and data analysts. An historical discussion and extensive bibliography on outliers may be obtained from Rider (1933), Grubbs (1950), and Anscombe (1960).

As pointed out by Dixon (1953), the existence of outliers actually presents two distinct problems:

(1) To identify any particular observation or observations that are "spurious", i.e., are from a population other than the one assumed to be under study.
(2) To obtain an analysis of the data which is not unduly affected by any spurious observations.

It is the second problem toward which this paper is directed. As a solution to this problem, a specific rule, which provides for the discarding or modifying of certain observations, may be used in an attempt to obtain a more "accurate" estimate of a desired parameter.

Anscombe (1960) emphasized that such a rule should be regarded in the same manner that a homeowner would regard a household fire insurance policy. A homeowner, accepting the fact that fires do occur, is concerned with buying an insurance policy which offers good protection for a moderate premium. Likewise, an experimental scientist, accepting the fact that spurious observations

Received March 1967; Revised Dec. 1967.

* Present Address: HRB-Singer, Inc., State College, Pa.

do occur, is concerned with using a rule which provides a relatively small mean square error (MSE), i.e., provides good protection, when a spurious observation is present. At the same time, the experimenter would want the rule not to seriously inflate the MSE if no spurious observation is present, i.e., the rule should have a moderate premium.

If we consider only rules which provide an unbiased estimate in the null case, we may define:

Premium = Fractional increase in variance due to using a given rule instead of the usual estimator, in the null case.
Protection = Fractional reduction in MSE due to using a given rule instead of the usual estimator, when a spurious observation is present.

Of course, if we apply a rule in the null case, Protection = −Premium.

The three basic assumptions underlying Anscombe's "Premium-Protection" approach are:

(i) The factors causing a spurious observation will not affect any other observation.
(ii) Computation costs and sampling costs may be ignored.
(iii) No prior information exists about unknown parameters or about which observation, if any, is spurious.

Most of the work which has been done on the "Premium-Protection" approach has been concerned with samples from the normal distribution when it is desired to estimate the mean μ. Although he also considered more general situations, Anscombe (1960) applied the "Premium-Protection" approach to investigate a rejection rule for outliers in samples of size three and four from a normal distribution. Further work was done for a sample size of three by Anscombe and Barron (1966), who considered a rejection rule and a modification rule. Veale and Huntsberger (1965) have also dealt with the outlier problem from the "Premium-Protection" approach when σ^2 is known, and used the value of the residual with largest magnitude to assign a weight to the corresponding observation for use in the estimation of μ. Gebhardt (1964) essentially used the "Premium-Protection" approach in a decision-theoretic framework. He assumed, however, that any spuriousity parameter was known, an assumption which he relaxed somewhat in a subsequent paper. (Gebhardt (1966)).

In this paper we shall, using the "Premium-Protection" approach, investigate three specific rules as applied to small samples from the normal distribution $N(\mu, \sigma^2)$ when it is desired to estimate the mean μ, and it is assumed that *at most* one spurious observation is present, either from $N(\mu + a\sigma, \sigma^2)$ or $N(\mu, (1 + b)\sigma^2)$. If the occurrence of a spurious observation is a "rare" event, the assumption that at most one appears in a small sample should not be too far removed from reality.

We label the rules we shall investigate as Anscombe's rule, the Winsorization rule, and the Semiwinsorization rule, which we abbreviate to the A-rule, the W-rule, and the S-rule, respectively. Each of these rules provides an unbiased estimate of the desired parameter μ in the null case, i.e., when no spurious observation is present.

Further, each of the rules discussed depends on a statistic exceeding a constant of the type $C(n, p, r)$ where n is the sample size, p is the premium to be paid, and r is the rule used. In our discussion of the rules, the generic notation C is used, but it should be held in mind that these constants are functions of (n, p, r).

We now summarize the notation that will be used in the subsequent sections. Let us consider a sample of n independent observations $(y_1, \cdots, y_n)$ where it is hoped that the underlying distribution from which this sample is taken is $N(\mu, \sigma^2)$. Denote the ordered observations by $y_{(1)} < \cdots < y_{(n)}$, define the residuals $z_i = y_i - \bar{y}$, $(i = 1, \cdots, n)$, where $\bar{y} = 1/n \sum_1^n y_i$, and the ordered residuals as $z_{(1)} < \cdots < z_{(n)}$. Let us further define the quantities:

$$\bar{y}_{(1)} = \frac{1}{n-1} \sum_2^n y_{(i)} \qquad \bar{y}_{(n)} = \frac{1}{n-1} \sum_1^{n-1} y_{(i)}$$

$$s^2 = \frac{1}{n-1} \sum_1^n (y_i - \bar{y})^2 \qquad \bar{y}_{(i,k)} = \frac{1}{n}\left[\sum_1^n y_{(i)} + y_{(j)} - y_{(k)} \right]$$

Anscombe's rule (the A-rule)

Anscombe (1960) proposed a rejection rule for guarding against a spurious observation when the mean μ is to be estimated. This rule is such that the suspect observation is discarded, and estimation proceeds using the remaining $(n - 1)$ observations as a "new" sample.

When σ^2 is known and μ is to be estimated, the A-rule uses the estimator

$$\hat{\mu}_A = \begin{cases} \bar{y} & \text{if } |z_{(1)}| < C\sigma \text{ and } |z_{(n)}| < C\sigma \\ \bar{y}_{(1)} & \text{if } |z_{(1)}| \geq C\sigma \text{ and } |z_{(1)}| > |z_{(n)}| \\ \bar{y}_{(n)} & \text{if } |z_{(n)}| \geq C\sigma \text{ and } |z_{(n)}| > |z_{(1)}| \end{cases}$$

When σ^2 is unknown and μ is to be estimated, the estimator is of the same form, with σ replaced by s throughout.

The Winsorization rule (The W-rule)

The term "Winsorization" has been applied to the procedure suggested by C. P. Winsor (see Dixon (1960)), whereby the value of a suspect observation is replaced by the value of the nearest retained observation, so that not all the information contained in the suspect observation is thrown out.

Applying Winsor's technique to construct a rule when σ^2 is known and μ is to be estimated, we arrive at an estimator which may be written as

$$\hat{\mu}_w = \begin{cases} \bar{y} & \text{if } |z_{(1)}| < C\sigma \text{ and } |z_{(n)}| < C\sigma \\ \bar{y}_{(2,1)} & \text{if } |z_{(1)}| \geq C\sigma \text{ and } |z_{(1)}| > |z_{(n)}| \\ \bar{y}_{(n-1,n)} & \text{if } |z_{(n)}| \geq C\sigma \text{ and } |z_{(n)}| > |z_{(1)}| \end{cases}$$

When σ^2 is unknown and μ is to be estimated, the estimator is of the same form with σ replaced by s.

The Semiwinsorization Rule (The S-rule)

The technique which we label "Semiwinsorization" is a modification of the Winsorization procedure. If any observation is deemed suspicious using the Semiwinsorization rule, the value of the statistic used by the rule is replaced by that of the appropriate boundary, with the resulting estimation subject to the constraint imposed by this procedure. The application of this rule will be made clearer by looking at the various cases under study.

For example, when σ^2 is known and μ is to be estimated, the S-rule uses estimator

$$\hat{\mu}_s = \begin{cases} \bar{y} & \text{if } |z_{(1)}| < C\sigma \text{ and } |z_{(n)}| < C\sigma \\ \frac{1}{n}[(n-1)\bar{y}_{(1)} + \bar{y} - C\sigma] & \text{if } |z_{(1)}| \geq C\sigma \text{ and } |z_{(1)}| > |z_{(n)}| \\ \frac{1}{n}[(n-1)\bar{y}_{(n)} + \bar{y} + C\sigma] & \text{if } |z_{(n)}| \geq C\sigma \text{ and } |z_{(n)}| > |z_{(1)}| \end{cases}$$

which is to say that if for instance, $z_{(1)}$ the smallest of the residuals, is such that $z_{(1)} \leq -C\sigma$ and $|z_{(1)}| > |z_{(n)}|$, then we put $z_{(1)}$ equal to the boundary, i.e., we have "$z_{(1)}$" $= -C\sigma$ so that "$y_{(1)} - \bar{y}$" $= -C\sigma$ or "$y_{(1)}$" $= \bar{y} - C\sigma$, and replace $y_{(1)}$ by "$y_{(1)}$" $= \bar{y} - C\sigma$ in the usual estimator of the mean, etc. When σ^2 is unknown and μ is to be estimated, the rule is of the same form, with σ replaced by s.

2. Functional Form of Premiums and Protections

The definitions of premium and protection may be rewritten using statistical notation. For a given rule which uses an estimator $\hat{\mu}$, we have:

$$\text{Premium} = [V(\hat{\mu}) - V(\bar{y})]/V(\bar{y}) \tag{2.1}$$

when all observations are from the assumed population $N(\mu, \sigma^2)$, and

$$\text{Protection} = [E(\bar{y} - \mu)^2 - E(\hat{\mu} - \mu)^2]/E(\bar{y} - \mu)^2 \tag{2.2}$$

when a spurious observation is present.

Premiums

When all n observations are from $N(\mu, \sigma^2)$, then $\bar{y} \sim N(\mu, \sigma^2/n)$, and hence $V(\bar{y}) = \sigma^2/n$. Thus, only $V(\hat{\mu})$ must be computed in order to evaluate the premium for a given rule.

To calculate the premium charged by the A-rule when the rejection region, i.e., a particular value of C, is given, we may rewrite the estimator $\hat{\mu}_A$ as $\hat{\mu}_A = \bar{y} + A$, where

$$A = A(\mathbf{z}) = A(z_1, \cdots, z_n) = \begin{cases} 0 & \text{if } |z_{(1)}| < C\sigma \text{ and } |z_{(n)}| < C\sigma \\ -z_{(1)}/(n-1) & \text{if } |z_{(1)}| \geq C\sigma \text{ and } |z_{(1)}| > |z_{(n)}| \\ -z_{(n)}/(n-1) & \text{if } |z_{(n)}| \geq C\sigma \text{ and } |z_{(n)}| > |z_{(1)}| \end{cases} \tag{2.3}$$

if σ^2 is known, with σ replaced by s if σ^2 is unknown. Since $\bar{y}$ is independent of the z_i's,

$$V(\hat{\mu}_A) = E(\bar{y} - \mu)^2 + E(A^2)$$
$$= \sigma^2/n + E(A^2)$$

Thus, from (2.1) we have for this rule that

$$\text{Premium} = n/\sigma^2 \cdot E(A^2) \tag{2.4}$$

Similarly, by rewriting the estimator $\hat{\mu}_w = \bar{y} + W$, where

$$W = W(\mathbf{z}) = W(z_1, \cdots, z_n)$$
$$= \begin{cases} 0 & \text{if } |z_{(1)}| < C\sigma \text{ and } |z_{(n)}| < C\sigma \\ [z_{(2)} - z_{(1)}]/n & \text{if } |z_{(1)}| \geq C\sigma \text{ and } |z_{(1)}| > |z_{(n)}| \\ [z_{(n-1)} - z_{(n)}]/n & \text{if } |z_{(n)}| \geq C\sigma \text{ and } |z_{(n)}| > |z_{(1)}| \end{cases} \tag{2.5}$$

if σ^2 is known, with σ replaced by s if σ^2 is unknown, we then have

$$V(\hat{\mu}_w) = (\sigma^2/n) + E(W^2)$$

and

$$\text{Premium} = (n/\sigma^2)E(W^2) \tag{2.6}$$

under the W-rule. Further, since the estimator $\hat{\mu}_s$ may be written as $\hat{\mu}_s = \bar{y} + S$, where

$$S = S(\mathbf{z}) = S(z_1, \cdots, z_n)$$
$$= \begin{cases} 0 & \text{if } |z_{(1)}| < C\sigma \text{ and } |z_{(n)}| < C\sigma \\ [-C\sigma - z_{(1)}]/n & \text{if } |z_{(1)}| \geq C\sigma \text{ and } |z_{(1)}| > |z_{(n)}| \\ [C\sigma - z_{(n)}]/n & \text{if } |z_{(n)}| \geq C\sigma \text{ and } |z_{(n)}| > |z_{(1)}| \end{cases} \tag{2.7}$$

if σ^2 is known, with σ replaced by s if σ^2 is unknown, it follows that

$$V(\hat{\mu}_s) = (\sigma^2/n) + E(S^2)$$

and

$$\text{Premium} = (n/\sigma^2) \cdot E(S^2) \tag{2.8}$$

Protections—Biased Mean Case

Let us assume now that $y_i \sim N(\mu, \sigma^2)$, $i = 1, \cdots, n-1$, that $y_n \sim N(\mu + a\sigma, \sigma^2)$, $-\infty < a < \infty$, and that the y_i's are independent. It will be noticed that the assumption that y_n is the biased observation does not incorporate any apriori knowledge into the rules, since the rules to be examined do not make use of any such assumption.

Hence, we have that

$$\bar{y} \sim N(\mu + a\sigma/n, \sigma^2/n)$$

and

$$E(\bar{y} - \mu)^2 = (\sigma^2/n)(1 + a^2/n)$$

Thus, to determine the protection provided by a given rule using estimator $\hat{\mu}$, only $E(\hat{\mu} - \mu)^2$ must be evaluated.

Again, in this case, $\bar{y}$ is independent of the z_i's, so

$$E(\hat{\mu}_A - \mu)^2 = E(\bar{y} - \mu - a\sigma/n)^2 + E(A + a\sigma/n)^2$$
$$= (\sigma^2/n) + E(A + a\sigma/n)^2$$

Thus, using the A-rule, we see from (2.2) that

$$\text{Protection} = -n^2E[A(A + 2a\sigma/n)]/\sigma^2(n + a^2). \tag{2.9}$$

In a similar manner, we have that using the W-rule,

$$\text{Protection} = -n^2E[W(W + 2a\sigma/n)]/\sigma^2(n + a^2). \tag{2.10}$$

Likewise, the S-rule has protection given by

$$\text{Protection} = -n^2E[S(S + 2a\sigma/n)]/\sigma^2(n + a^2). \tag{2.11}$$

Protections—Biased Variance Case

Let us assume that $y_i \sim N(\mu, \sigma^2)$ $i = 1, \cdots, n - 1$, that $y_n \sim N(\mu, (1 + b)\sigma^2)$, $0 \leq b$, and that y_i's are independent. Again we note that identifying y_n as the biased observation does not bring any apriori knowledge into the rules.

Under these assumptions,

$$\bar{y} \sim N\left(\mu, \frac{(n + b)}{n^2}\sigma^2\right)$$

and thus,

$$E(\bar{y} - \mu)^2 = \frac{(n + b)}{n^2}\sigma^2$$

Hence, we need only evaluate $E(\hat{\mu} - \mu)^2$ to determine the protection. For the A-rule we have

$$E(\hat{\mu}_A - \mu)^2 = E(\bar{y} - \mu)^2 + 2E[A(\bar{y} - \mu)] + E(A^2)$$

In this case, $\bar{y}$ is *not* independent of the z_i's, and we are left with

$$E(\hat{\mu}_A - \mu)^2 = \frac{(n + b)}{n^2}\sigma^2 + E[A\{2(\bar{y} - \mu) + A\}]$$

where the expectation is over the z_i's and $\bar{y}$. Thus, the protection afforded by the A-rule is

$$\text{Protection} = \frac{-n^2E[A\{2(\bar{y} - \mu) + A\}]}{\sigma^2(n + b)} \tag{2.12}$$

Similarly, for the W-rule, we have

$$\text{Protection} = \frac{-n^2E[W\{2(\bar{y} - \mu) + W\}]}{\sigma^2(n + b)} \tag{2.13}$$

and for the S-rule,

$$\text{Protection} = \frac{-n^2E[S\{2(\bar{y} - \mu) + S\}]}{\sigma^2(n + b)} \tag{2.14}$$

3. Analysis When σ^2 is Known

The case $n = 3$

To obtain the constant C corresponding to a given rule and premium, we must, of course, investigate the rule in the null case. Thus, our sample consists of three independent observations y_1, y_2, y_3 each from $N(\mu, \sigma^2)$ where μ is to be estimated. Without loss of generality we may take $\sigma^2 = 1$ if σ^2 is assumed known.

With $\sigma^2 = 1$, the joint density of $(z_{(1)}, z_{(3)})$ in the null case takes the form,

$$f(z_{(1)}, z_{(3)}) = (3\sqrt{3}/\pi) \exp\{-(z_{(1)}^2 + z_{(1)}z_{(3)} + z_{(3)}^2)\} \tag{3.1}$$

over the region

$$\left\{(z_{(1)}, z_{(3)}) \;\middle|\; \frac{-z_{(1)}}{2} < z_{(3)} < -2z_{(1)}\right\}.$$

In his 1960 paper, Anscombe calculated the premium corresponding to a given rejection constant for a sample of three, essentially by finding the distribution of z_M, the residual with largest magnitude. (It should be noted that although the density of $z_{(n)}$, or equivalently $z_{(1)}$, was given by McKay (1935) and Nair (1948) for general n, this is not sufficient to determine the density of z_M, since z_M depends on both $z_{(1)}$ and $z_{(n)}$). However, in order to investigate the W-rule, the distribution of z_M does not suffice, and the joint density of $(z_{(1)}, z_{(3)})$ is required. For this reason, we have approached the computational problem by using this density.

For $n = 3$ and $\sigma^2 = 1$, we have from (2.3) and (2.4) that for the A-rule,

$$\begin{aligned}\text{Premium} &= 3E(A^2) \\ &= 3\int_{G_1} \frac{z_{(1)}^2}{4} f(z_{(1)}, z_{(3)})\, dz_{(3)}\, dz_{(1)} \\ &\quad + 3\int_{G_2} \frac{z_{(3)}^2}{4} f(z_{(1)}, z_{(3)})\, dz_{(1)}\, dz_{(3)} \\ &= \frac{3}{2}\int_{G_1} z_{(1)}^2 f(z_{(1)}, z_{(3)})\, dz_{(3)}\, dz_{(1)}\end{aligned} \tag{3.2}$$

where

$$G_1 = \left\{(z_{(1)}, z_{(3)}) \mid -\infty < z_{(1)} < -C, \frac{-z_{(1)}}{2} < z_{(3)} < -z_{(1)}\right\}$$

and

$$G_2 = \left\{(z_{(1)}, z_{(3)}) \mid C < z_{(3)} < \infty, -z_{(3)} < z_{(1)} < -\frac{z_{(3)}}{2}\right\}.$$

For a given premium of p, then, the constant C must be found which when substituted into (3.2) makes the expression equal to p. Expressions similar to (3.2) may be derived for the W-rule from (2.5) and (2.6) and for the S-rule from (2.7) and (2.8).

Having chosen a rule and a premium with corresponding constant C, we may

calculate protection in the biased mean case, i.e., when $y_3 \sim N(\mu + a\sigma, \sigma^2)$, using the formulas of section 2. In order to do this, we need the joint density of $(z_{(1)}, z_{(3)})$, which now reflects the presence of the biased observation. Let us denote this density by $g(z_{(1)}, z_{(3)}, a)$, where "a" is the bias. If $\sigma^2 = 1$, this density is given by

$$
\begin{aligned}
&g(z_{(1)}, z_{(3)}; a) \\
&= \frac{\sqrt{3}}{\pi} \exp\left\{-\left[\left(z_{(1)} + \frac{a}{3}\right)^2 + \left(z_{(1)} + \frac{a}{3}\right)\left(z_{(3)} + \frac{a}{3}\right) + \left(z_{(3)} + \frac{a}{3}\right)^2\right]\right\} \\
&\quad + \frac{\sqrt{3}}{\pi} \exp\left\{-\left[\left(z_{(1)} - \frac{2a}{3}\right)^2 + \left(z_{(1)} - \frac{2a}{3}\right)\left(z_{(3)} + \frac{a}{3}\right) + \left(z_{(3)} + \frac{a}{3}\right)^2\right]\right\} \\
&\quad + \frac{\sqrt{3}}{\pi} \exp\left\{-\left[\left(z_{(1)} + \frac{a}{3}\right)^2 + \left(z_{(1)} + \frac{a}{3}\right)\left(z_{(3)} - \frac{2a}{3}\right) + \left(z_{(3)} - \frac{2a}{3}\right)^2\right]\right\} \qquad (3.3)
\end{aligned}
$$

over the region

$$
\left\{(z_{(1)}, z_{(3)}) \;\middle|\; \frac{-z_{(1)}}{2} < z_{(3)} < -2z_{(1)}\right\}.
$$

We may also calculate protection in the biased variance case, i.e., when $y_3 \sim N(\mu, (1 + b)\sigma^2)$, $0 \leq b$.

If $\sigma^2 = 1$, the joint density of $(z_{(1)}, z_{(3)}, \bar{y})$ is, for a bias of "b",

$$
\begin{aligned}
&h(z_{(1)}, z_{(3)}, \bar{y}; b) \\
&= 6(2\pi)^{-\frac{1}{2}}(1 + b)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\frac{2 + b}{1 + b}(z_{(1)}^2 + z_{(3)}^2)\right.\right. \\
&\quad \left.\left. + \frac{2}{1 + b} z_{(1)}z_{(3)} + \frac{2b}{1 + b}(\bar{y} - \mu)(z_{(1)} + z_{(3)}) + \frac{3 + 2b}{1 + b}(\bar{y} - \mu)^2\right)\right\} \\
&\quad + 6(2\pi)^{-\frac{1}{2}}(1 + b)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(2z_{(1)}^2 + \frac{2 + b}{1 + b} z_{(3)}^2\right.\right. \\
&\quad \left.\left. + 2z_{(1)}z_{(3)} - \frac{2b}{1 + b} z_{(3)}(\bar{y} - \mu) + \frac{3 + 2b}{1 + b}(\bar{y} - \mu)^2\right)\right\} \\
&\quad + 6(2\pi)^{-\frac{1}{2}}(1 + b)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(2z_{(3)}^2 + \frac{2 + b}{1 + b} z_{(1)}^2\right.\right. \\
&\quad \left.\left. + 2z_{(1)}z_{(3)} - \frac{2b}{1 + b} z_{(1)}(\bar{y} - \mu) + \frac{3 + 2b}{1 + b}(\bar{y} - \mu)^2\right)\right\} \qquad (3.4)
\end{aligned}
$$

over the region

$$
\left\{(z_{(1)}, z_{(3)}, \bar{y}) \mid -\infty < \bar{y} < \infty, -\infty < z_{(1)} < 0, \frac{-z_{(1)}}{2} < z_{(3)} < -2z_{(1)}\right\}
$$

These are the densities used to compute the expectations required to obtain the protections afforded by a specific rule for a given bias. The resulting integrals can be reduced to at most double integrals which may be easily evaluated by numerical integration.

The case $n > 3$

For a sample of n observations, the formulas for protections and premiums involve, in general, $(n - 1)$-fold integrals, which, unfortunately, cannot be reduced. For $n > 3$, then, Monte Carlo procedures are utilized.

The method used to determine the constant C for a given premium, sample size, and rule consists of sampling over certain regions a given number of times, say N, by means of a specific sampling technique. The theory underlying this sampling is equivalent to that used in computing protections (discussed below), except that the inherent symmetry of the null case is taken into consideration to simplify the computer programs.

The actual value of N used to determine the value of C varied for each rule, premium, and sample size, and was such that reasonably small standard errors of the estimated premiums were obtained. In practice moderate values of N were used to establish rough bounds on C, and larger values of N were used for iteration to determine C to two decimal places.

To deal with protections in the biased mean case more easily, we may reformulate the problem as follows:

Let $y_1, \cdots, y_n$ be independently distributed, where $y_i \sim N(0, 1)$ for $i = 1, \cdots, n - 1$, and $y_n \sim N(a, 1)$.

Thus, this restatement reduces the original problem to one of "estimating" the mean $\mu = 0$, where $\sigma^2 = 1$. This of course, does not alter the results, since σ^2 is assumed known, and the rejection rules make no use of any information about μ.

The joint density of $(z_1, \cdots, z_{n-1})$ is, with $\mu = 0$ and $\sigma^2 = 1$,

$$g_1(z_1, \cdots, z_{n-1}; a) = \frac{n^{\frac{1}{2}}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\left[2\sum_1^{n-1} z_i^2 + 2a\sum_1^{n-1} z + 2\sum_{i<j}^{n-1} z_i z_j + \frac{n-1}{n}a^2\right]\right\}$$

$$-\infty < z_i < \infty, \qquad i = 1, 2, \cdots, n-1.$$

That is, the vector $\mathbf{z}' = (z_1, \cdots, z_{n-1})$ is multivariate normal with mean vector $-(a/n)(1, \cdots, 1)$ and variance-covariance matrix $\mathbf{I}_{n-1} - (1/n)\mathbf{1}\mathbf{1}' =$

$$\mathbf{I}_{n-1} - \frac{1}{n}\begin{bmatrix} 1 & 1 & \cdot & \cdot & 1 & 1 \\ 1 & 1 & \cdot & \cdot & 1 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & 1 & 1 \\ 1 & 1 & \cdot & \cdot & 1 & 1 \end{bmatrix}_{n-1}$$

where $\mathbf{I}_{n-1}$ is the identity matrix.

Let us now consider the transformation

$$\mathbf{w} = \mathbf{L}\mathbf{z} \tag{3.5}$$

where $\mathbf{w}' = (w_1, \cdots, w_{n-1})$ and where

$$\mathbf{L} = \begin{bmatrix} \sqrt{\frac{n}{n-1}} & \sqrt{\frac{n}{n-1}} & \sqrt{\frac{n}{n-1}} & \cdots & \sqrt{\frac{n}{n-1}} \\ \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} & 0 & \cdots & 0 \\ \sqrt{\frac{1}{6}} & \sqrt{\frac{1}{6}} & -2\sqrt{\frac{1}{6}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots\ \vdots\ \vdots & \vdots \\ \sqrt{\frac{1}{(n-2)(n-1)}} & \sqrt{\frac{1}{(n-2)(n-1)}} & \sqrt{\frac{1}{(n-1)(n-1)}} & \cdots & \frac{-(n-2)}{\sqrt{(n-2)(n-1)}} \end{bmatrix}_{n-1}$$

Thus, $\mathbf{w}'$ is multivariate normal with mean $-a\sqrt{(n-1/n)}(1, 0, \cdots, 0)$ and variance-covariance matrix $\mathbf{I}_{n-1}$. Note that, by definition,

$$z_n = -\sum_{1}^{n-1} z_i = -w_1\sqrt{\frac{n-1}{n}}.$$

The inverse of transformation (3.5) is

$$\mathbf{z} = \mathbf{L}^{-1}\mathbf{w} \tag{3.6}$$

where $\mathbf{L}^{-1}$ is obtained by replacing the first row of $\mathbf{L}$ by the vector $1/\sqrt{n(n-1)}$ $(1, 1, \cdots, 1)$ and then transposing the resulting matrix. Heace,

$$z_{n-1} = \frac{w_1}{\sqrt{n(n-1)}} - w_{n-1}\sqrt{\frac{n-2}{n-1}}.$$

To calculate the protection given by the A-Rule, the W-Rule, and the S-Rule, we need $E[A(A + 2a/n)]$, $E[W(W + 2a/n)]$, and $E[S(S + 2a/n)]$, respectively, where A, W, and S are the functions of $\mathbf{z}$ defined previously. We shall describe the procedure used to compute the required expectation $E[A(A + 2a/n)]$. The expectations needed for evaluation of protection given by the other two rules in the biased mean case may be computed by analogous procedures. The procedure when dealing with the biased variance case will not be considered here, since it is similar, although greater use may be made of symmetry conditions. The details are given in Smith (1966).

In order to simplify our derivations, let us denote the residual with largest magnitude by z_M. Note that the random variable A (see (2.3)) may be written as

$$A = \begin{cases} 0 & \text{if } |z_M| < C \\ \dfrac{-z_M}{n-1} & \text{if } |z_M| \geq C \end{cases} \tag{3.7}$$

Rewriting, we have

$$E\left[A\left(A + \frac{2a}{n}\right)\right] = \text{Prob}\,(|z_M| < C)E\left[\frac{2a}{n}A + A^2 \,\middle|\, |z_M| < C\right] + \text{Prob}\,(|z_M| \geq C)E\left[\frac{2a}{n}A + A^2 \,\middle|\, |z_M| \geq C\right].$$

Using (3.7), we may write

$$E\left[A\left(A+\frac{2a}{n}\right)\right] = \text{Prob}\,(|z_M| \geq C)E\left[\frac{-2az_M}{n(n-1)} + \frac{z_M^2}{(n-1)^2} \,\middle|\, |z_M| \geq C\right]$$

$$= \sum_{i=1}^{n} \text{Prob}\,(|z_i| \geq C \quad \text{and} \quad M = i)$$

$$\cdot E\left[\frac{-2az_i}{n(n-1)} + \frac{z_i^2}{(n-1)^2} \,\middle|\, |z_i| \geq C \quad \text{and} \quad M = i\right]$$

Now, for $i = 1, \cdots, n-1$, the symmetry of the situation gives

$$\text{Prob}\,(|z_i| \geq C \quad \text{and} \quad M = i)E\left[\frac{-2az_i}{n(n-1)} + \frac{z_i^2}{(n-1)^2} \,\middle|\, |z_i| \geq C \quad \text{and} \quad M = i\right]$$

$$= \text{Prob}\,(|z_{n-1}| \geq C \quad \text{and} \quad M = n-1)$$

$$\cdot E\left[\frac{-2az_{n-1}}{n(n-1)} + \frac{z_{n-1}^2}{(n-1)^2} \,\middle|\, |z_{n-1}| \geq C \quad \text{and} \quad M = n-1\right]$$

Considering $\mathbf{z}$-space, and defining the regions

$$U^* = \{\mathbf{z} \mid |z_{n-1}| \geq C \quad \text{and} \quad M = n-1\}$$

$$V^* = \{\mathbf{z} \mid |z_n| \geq C \quad \text{and} \quad M = n\}$$

where $z_n = -\sum_1^{n-1} z_i$, we have

$$E\left[A\left(A+\frac{2a}{n}\right)\right] = (n-1)\,\text{Prob}\,(\mathbf{z} \in U^*)E\left[\frac{-2az_{n-1}}{n(n-1)} + \frac{z_{n-1}^2}{(n-1)^2} \,\middle|\, \mathbf{z} \in U^*\right]$$

$$+ \text{Prob}\,(\mathbf{z} \in V^*)E\left[\frac{-2az_n}{n(n-1)} + \frac{z_n^2}{(n-1)^2} \,\middle|\, \mathbf{z} \in V^*\right] \tag{3.8}$$

The procedure which we have constructed involves taking a number, N_1, of random samples in $\mathbf{z}$-space when $\mathbf{z} \in U^*$, and N_2 when $\mathbf{z} \in V^*$.
If we define the function

$$g(z; a) = \begin{cases} \dfrac{-2az}{n(n-1)} + \dfrac{z^2}{(n-1)^2} & \text{if} \quad \mathbf{z} \in U^* \\ 0 & \text{if} \quad \mathbf{z} \in U - U^* \end{cases} \tag{3.9}$$

where

$$U = \{\mathbf{z} \mid |z_{n-1}| \geq C \quad \text{and} \quad |z_{n-1}| \geq |z_n|\},$$

we may write

$$\text{Prob}\,(z \in U^*)E\left[\frac{-2az_{n-1}}{n(n-1)} + \frac{z_{n-1}^2}{(n-1)^2} \,\middle|\, \mathbf{z} \in U^*\right]$$

$$= \text{Prob}\,(z \in U)E[g(z_{n-1}\,; a) \mid \mathbf{z} \in U] \tag{3.10}$$

In terms of the w's (3.5), $\mathbf{z} \in U$ is equivalent to

$$\mathbf{w} \in \left\{\mathbf{w} \,\middle|\, \left|\frac{w_1}{\sqrt{n(n-1)}} - w_{n-1}\sqrt{\frac{n-2}{n-1}}\right| \geq \max\left(C, \left|w_1\sqrt{\frac{n-1}{n}}\right|\right)\right\}$$

Now, it is easily seen that by using the orthogonal transformation

$$\begin{aligned} x_1 &= \frac{-\sqrt{n(n-2)}}{n-1}\left(w_1 + a\sqrt{\frac{n-1}{n}}\right) - \frac{w_{n-1}}{n-1} \\ x_2 &= \frac{1}{n-1}\left(w_1 + a\sqrt{\frac{n-1}{n}}\right) - \frac{\sqrt{n(n-2)}}{n-1}w_{n-1} \\ x_3 &= w_2 \\ x_4 &= w_3 \\ &\vdots \\ x_{n-1} &= w_{n-2} \end{aligned} \tag{3.11}$$

we have that $\mathbf{x}' = (x_1, \cdots, x_{n-1})$ is distributed as a spherical $(n-1)$-variate normal random variable, i.e., with mean $\mathbf{0}$ and variance-covariance matrix $\mathbf{I}_{n-1}$.

Thus, $\mathbf{z} \in U$ is equivalent to $\mathbf{x} \in R$, where

$$R = \left\{\mathrm{x} \;\middle|\; \left|x_2\sqrt{\frac{n-1}{n}} - \frac{a}{n}\right| \geq \max\left(C, \left|\frac{x_2}{\sqrt{n(n-1)}} - x_1\sqrt{\frac{n-2}{n-1}} - \frac{a(n-1)}{n}\right|\right)\right\}$$

The region R is the union of two disjoint regions R_1 and R_2, as Figure 1 indicates.

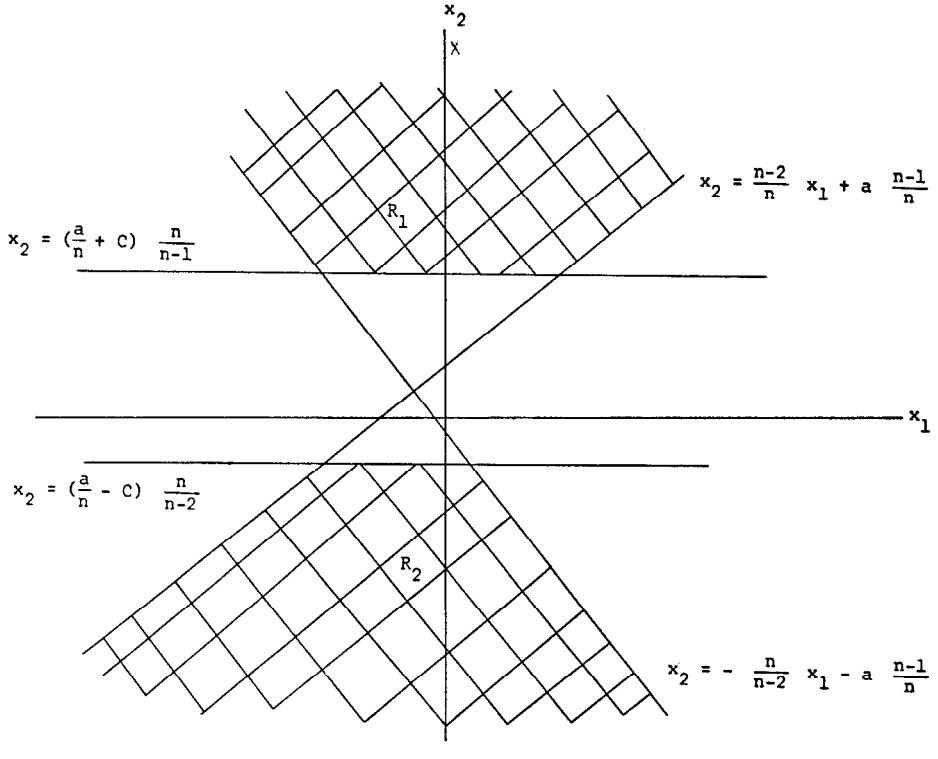

FIGURE 1—The region $R_1 \cup R_2$ in the (x_1, x_2)-plane.

Let us now define $P_1 = \text{Prob}\ (\mathbf{x} \in R_1)$ and $P_2 = \text{Prob}\ (\mathbf{x} \in R_2)$, which may be easily found with the aid of the computer. Note that

$$\text{Prob}\ (\mathbf{z} \in U) = \text{Prob}\ (\mathbf{x} \in R) = P_1 + P_2$$

Now, if we wish to select, at random, $\mathbf{x} \in R$, we must weight the region R_1, with probability $p = p_1/p_1 + p_2$, and the region R_2 with probability $(1 - p)$. The problem then becomes one of selecting $\mathbf{x} \in R_1$ (or R_2) according to the probability law given by the $(n - 1)$-variate spherical normal density.

Since there exist computer routines which generate pseudo-random normal deviates, we may easily obtain $x_3, \cdots, x_{n-1}$. However, the restrictions imposed by the region under consideration present difficulties in the generation of x_1 and x_2, since (x_1, x_2), the points of interest, fall in the tails of the underlying probability distribution.

To circumvent these difficulties, we follow Box and Muller (1958) and consider a sample of two, say (u_1, u_2), from a uniform distribution on (0, 1). Thus, the joint density of (u_1, u_2) is

$$f(u_1, u_2) = 1 \qquad \begin{matrix} 0 < u_1 < 1 \\ 0 < u_2 < 1 \end{matrix}$$

Now, the transformation

$$\begin{aligned} u_1 &= \exp\left[-\tfrac{1}{2}(x_1^2 + x_2^2)\right] \\ u_2 &= \frac{1}{2\pi}\tan^{-1}\left(\frac{x_2}{x_1}\right) \end{aligned} \tag{3.12}$$

is one-to-one, with Jacobian

$$-\frac{1}{2\pi}\exp\left[-\tfrac{1}{2}(x_1^2 + x_2^2)\right]$$

Hence, the joint density of (x_1, x_2) is

$$p(x_1, x_2) = \frac{1}{2\pi}\exp\left[-\tfrac{1}{2}(x_1^2 + x_2^2)\right] \qquad \begin{matrix} -\infty < x_1 < \infty \\ -\infty < x_2 < \infty \end{matrix}$$

Suppose now that (x_1, x_2) is to be chosen at random, given that $\mathbf{x} \in R_2$. (In the following discussion we shall concern ourselves only with $\mathbf{x} \in R_2$, although similar results hold for $\mathbf{x} \in R_1$,) Due to symmetry, we need only consider biases, "a", which are positive. For ease of notation, let $K = (a/n - C)\sqrt{n/n - 1}$. Since the inverse of transformation (3.12) is

$$\begin{aligned} x_1 &= (-2 \ln u_1)^{\frac{1}{2}} \cos(2\pi u_2) \\ x_2 &= (-2 \ln u_1)^{\frac{1}{2}} \sin(2\pi u_2) \end{aligned} \tag{3.13}$$

we have that

$$-\infty < x_2 \leq K \Rightarrow -\infty < (-2 \ln u_1)^{\frac{1}{2}} \sin(2\pi u_2) \leq K$$

With $0 \leq a \leq C$ and $K < 0$, it follows that $\frac{1}{2} < u_2 < 1$, and thus we have

$$0 < u_1 < \exp\left\{-\frac{1}{2}\left[\frac{K}{\sin(2\pi u_2)}\right]^2\right\}$$

Now, the maximum value which u_1 can attain is exp $(-\frac{1}{2}K^2)$, so we may obtain a pseudo-random number, say u_1^*, on (0, 1), multiply it by exp $(-\frac{1}{2}K^2)$ to obtain a random number u_1 less than this maximum value. Similarly, we may obtain at random, $u_2 \in (\frac{1}{2}, 1)$.

We may use this pair (u_1, u_2), and apply the transformation of (3.13) to give us an x_2 which has a high probability of being less than K. If $x_2 > K$, we repeat the above steps, while if $x_2 \leq K$, we proceed to obtain x_1, not from transformation (3.13) which would impose an unwanted restriction, but by means of the pseudo-random normal deviate generator mentioned above. Having obtained (x_1, x_2), we check to see if the resultant $\mathbf{x} \in R_2$. If not, the above procedure is repeated in its entirety until a satisfactory pair (x_1, x_2) is found.

The procedure outlined above is quite efficient when Prob $(\mathbf{x} \in R_2 \mid x_2 \leq K)$ is relatively large, as it is when $0 \leq a \leq C$. When this probability is small, as it is when $0 \leq C \leq a$, the difficulty is overcome by rotating the x_1 and x_2 axes to an orientation which allows a similar procedure to be used.

Assuming, then, that we have obtained $\mathbf{x} \in R$, we may, by means of the inverse of transformation (3.11), obtain $\mathbf{w}$ from the given $\mathbf{x}$. The z's are obtained via transformation (3.6), that is, $\mathbf{z} = \mathbf{L}^{-1}\mathbf{w}$. Thus, the z's have been chosen such that $\mathbf{z} \in U$. We then compute $g(z_{n-1}\,; a)$, where the function g is defined by (3.9), and keep track of how many times $\mathbf{z} \in U^*$, stopping when a total of N_1 repeats have been taken in U^*.

We will then have obtained a set of $\mathbf{z}$'s, N_1 of which are in U^*, and the rest of which, say T_1, are in $U - U^*$. If we denote the value of the function g for the k-th $\mathbf{z}$ by g_k, and recall that Prob $(\mathbf{z} \in U) = p_1 + p_2$, we may write (3.10) as

$$\text{Prob } (\mathbf{z} \in U^*)E\left[\frac{-2az_{n-1}}{n(n-1)} + \frac{z_{n-1}^2}{(n-1)^2} \,\middle|\, \mathbf{z} \in U^*\right] = \frac{p_1 + p_2}{N_1 + T_1} \sum_1^{N_1+T_1} g_k \,.$$

As may be seen from (3.8), this quantity provides that portion of $E[A(A + 2a/n)]$ involving U^*. An analogous procedure may be followed to compute the remaining portion which involves V^*.

4. Analysis When σ^2 is Unknown

The case $n = 3$

From section 1 and 2 we see that when σ^2 is unknown, the rules under consideration use estimators whose values are dependent upon the boundary

$$\max\,(|z_{(1)}|, |z_{(3)}|) = Cs$$

where

$$s^2 = \tfrac{1}{2}\sum_1^3 (y_i - \bar{y})^2 = \tfrac{1}{2}\sum_1^3 z_i^2 \,.$$

The ordered residuals are $z_{(1)} < z_{(2)} < z_{(3)}$, and since $z_{(2)} = -z_{(1)} - z_{(3)}$, we have that

$$s^2 = z_{(1)}^2 + z_{(1)}z_{(3)} + z_{(3)}^2$$

Let us now assume for the moment that $|z_{(1)}| > |z_{(3)}|$. Thus, our interest is in

the boundary $|z_{(1)}| = Cs$. Since we have assumed that $|z_{(1)}| > |z_{(3)}|$, i.e., $-z_{(1)} > z_{(3)}$, it follows that $|z_{(1)}| > s$. The region of definition of $(z_{(1)}, z_{(3)})$ is

$$\left\{(z_{(1)}, z_{(3)}) \;\middle|\; \frac{-z_{(1)}}{2} < z_{(3)} < -2z_{(1)}\right\}.$$

Thus, if $|z_{(1)}| > |z_{(3)}|$, then $s < |z_{(1)}| < (2/3^{\frac{1}{2}})s$, and likewise, if $|z_{(3)}| > |z_{(1)}|$, then $s < |z_{(3)}| < (2/3^{\frac{1}{2}})s$. Hence, if $C < 1$, a rejection (or modification) will always be made, and if $C > (2/3^{\frac{1}{2}})$, a rejection (or modification) will never be made.

Let us return to the assumption that $|z_{(1)}| > |z_{(3)}|$, and consider again the boundary $|z_{(1)}| = Cs$. Solving for $z_{(3)}$ in terms of $z_{(1)}$, we have that

$$z_{(3)} = z_{(1)}\left[-\tfrac{1}{2} \pm \frac{(4 - 3C^2)^{\frac{1}{2}}}{2C}\right].$$

This, coupled with the region of definition and the assumption that $|z_{(1)}| > |z_{(3)}|$, implies that the rejection (or modification) region for the observation $y_{(1)}$ corresponding to $z_{(1)}$ is

$$-z_{(1)} \cdot \max\left[\tfrac{1}{2}, \tfrac{1}{2} - \frac{(4 - 3C^2)^{\frac{1}{2}}}{2C}\right] < z_{(3)} < -z_{(1)} \cdot \min\left[1, \tfrac{1}{2} + \frac{(4 - 3C^2)^{\frac{1}{2}}}{2C}\right]$$

and since $(4 - 3C^2)^{\frac{1}{2}}/2C > 0$, the rejection (or modification) region is

$$\frac{-z_{(1)}}{2} < z_{(3)} < -Rz_{(1)}$$

or, equivalently,

$$-2z_{(3)} < z_{(1)} < \frac{-z_{(3)}}{R}$$

where

$$R = \begin{cases} \tfrac{1}{2} + \dfrac{(4 - 3C^2)^{\frac{1}{2}}}{2C} & \text{if } C \geq 1 \\ 1 & \text{if } C < 1 \end{cases}$$

Similarly, we find that when $|z_{(3)}| > |z_{(1)}|$, the rejection or modification region is

$$\frac{-z_{(1)}}{R} < z_{(3)} < -2z_{(1)}.$$

Since we have written the rejection (or modification) region in terms of $z_{(1)}$ and $z_{(3)}$, we need not be concerned with the distribution of s. Let us now define, for a given R, the events (or regions)

$$T_0 = \left\{(z_{(1)}, z_{(3)}) \mid 0 < z_{(3)} < \infty, \frac{-z_{(3)}}{R} < z_{(1)} < -z_{(3)}, -z_{(1)} < z_{(3)} < \frac{-z_{(1)}}{R}\right\}$$

$$T_1 = \left\{(z_{(1)}, z_{(3)}) \mid 0 < z_{(3)} < \infty, -2z_{(3)} < z_{(1)} < \frac{-z_{(3)}}{R}\right\}$$

$$T_3 = \left\{(z_{(1)}, z_{(3)}) \mid 0 < z_{(3)} < \infty, \frac{-z_{(1)}}{R} < z_{(3)} < -2z_{(1)}\right\}$$

Hence, using the definitions of T_0, T_1, and T_3, the estimators $\hat{\mu}_A$, $\hat{\mu}_W$, and $\hat{\mu}_S$, may be written as $\hat{\mu}_A = \bar{y} + A$, $\hat{\mu}_W = \bar{y} + W$, and $\hat{\mu}_S = \bar{y} + S$, where

$$A = \begin{cases} 0 & \text{if } (z_{(1)}, z_{(3)}) \in T_0 \\ \dfrac{-z_{(1)}}{2} & \text{if } (z_{(1)}, z_{(3)}) \in T_1 \\ \dfrac{-z_{(3)}}{2} & \text{if } (z_{(1)}, z_{(3)}) \in T_3 \end{cases} \tag{4.1}$$

$$W = \begin{cases} 0 & \text{if } (z_{(1)}, z_{(3)}) \in T_0 \\ -\dfrac{z_{(3)} + 2z_{(1)}}{3} & \text{if } (z_{(1)}, z_{(3)}) \in T_1 \\ -\dfrac{z_{(1)} + 2z_{(3)}}{3} & \text{if } (z_{(1)}, z_{(3)}) \in T_3 \end{cases} \tag{4.2}$$

and

$$S = \begin{cases} 0 & \text{if } (z_{(1)}, z_{(3)}) \in T_0 \\ -\dfrac{Cv + z_{(1)}}{3} & \text{if } (z_{(1)}, z_{(3)}) \in T_1 \\ -\dfrac{Cv - z_{(3)}}{3} & \text{if } (z_{(1)}, z_{(3)}) \in T_3 \end{cases} \tag{4.3}$$

For a given constant C, the premiums are of the same functional form as in the case when σ^2 is known. Hence, we still need to evaluate the expectations $E(A^2)$, $E(W^2)$, and $E(S^2)$, and substitute the values of these quantities into (2.4), (2.6), and (2.8), respectively, to obtain the premiums "charged" by the three rules under consideration. In order to do this, the density of $(z_{(1)}, z_{(3)})$ is required. It is easily verified that this density is $(1/\sigma^2)f(z_{(1)}/\sigma, z_{(3)}/\sigma)$, where f is given by (3.1).

Similar results hold for protections and the densities of $(z_{(1)}, z_{(3)})$ in the cases where a spurious observation is present. All required quantities may be obtained by means of numerical integration.

The case $n > 3$

The theoretical aspects of computing premiums and protections when σ^2 is unknown present no extreme difficulties when a sample size n larger than three is considered. However, the practical problem of deriving a reasonable method of obtaining numerical results has been severe because of the computer time that would be needed to evaluate the required $(n-1)$-fold integrals by numerical integration, while consideration of Monte Carlo procedures get bogged down because of the form of the regions in which samples must be taken. For example, sampling of $(z_1, \cdots, z_{n-1})$ is required subject to conditions of the form

$$z_{n-1}^2 \geq \frac{2C^2}{n-1}(z_1^2 + \cdots + z_{n-1}^2 + z_1z_2 + \cdots + z_{n-1}z_{n-1}).$$

The products z_iz_j add considerable complications to the problem, for essentially we need to sample in a cone in $(n-1)$-space, which is no easy task.

5. Discussion of Results

From the following tables and graphs it can be seen that, in general, for estimation of μ with σ^2 known when a spurious observation from $N(\mu + a\sigma, \sigma^2)$ is present, the S-rule is best for small values of a, the W-rule is best for moderate values of a, and the A-rule is best for large values of a. When a spurious observation from $N(\mu, (1 + b)\sigma^2)$ is present, the S-rule performs best for small

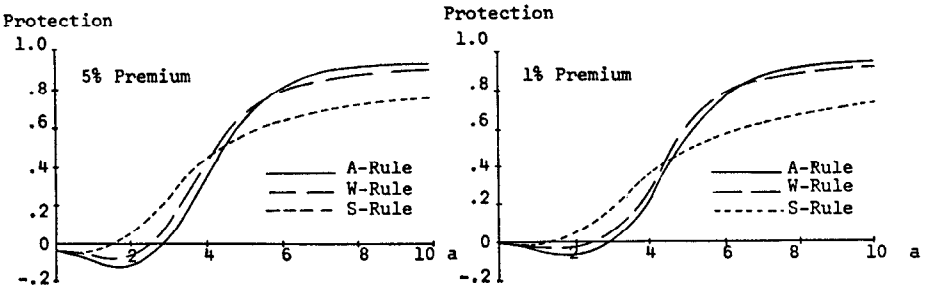

FIGURE 2—Protections corresponding to premiums of 5% and 1% when a spurious observation from $N(\mu + a\sigma, \sigma^2)$ is present in a sample of size three and σ^2 is known. (Symmetric about $a = 0$)

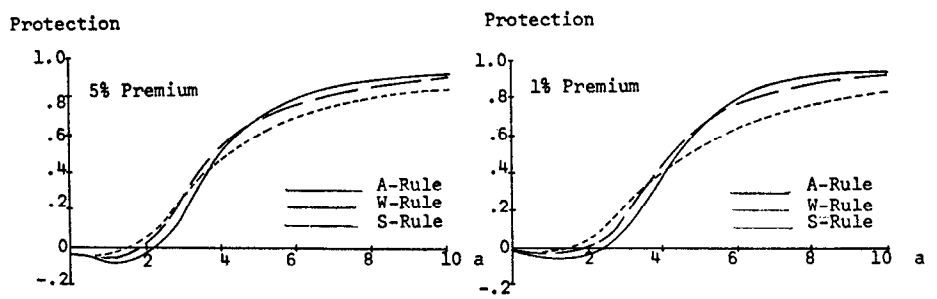

FIGURE 3—Protections corresponding to premiums of 5% and 1% when a spurious observation from $N(\mu + a\sigma, \sigma^2)$ is present in a sample of size six and σ^2 is known. (Symmetric about $a = 0$)

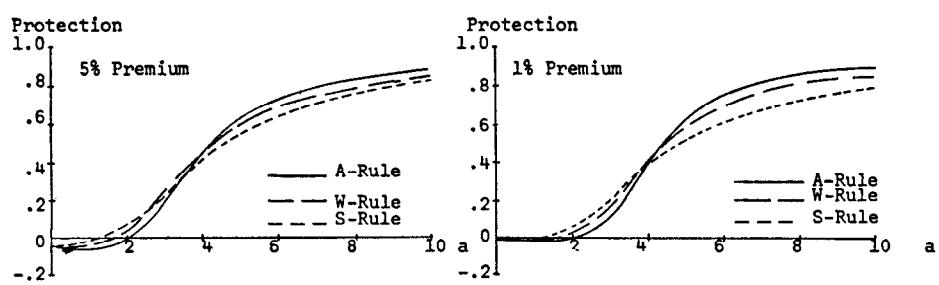

FIGURE 4—Protections corresponding to premiums of 5% and 1% when a spurious observation from $N(\mu + a\sigma, \sigma^2)$ is present in a sample of size ten and σ^2 is known. (Symmetric about $a = 0$)

values of b, while the W-rule performs best for large values of b. In this case there is no interval over which the A-rule performs best.*

It seems to us that we may look at our results from two viewpoints:

(1) If we cling to our original concept of no prior knowledge regarding the magnitude of bias of a spurious observation if any should occur, we have

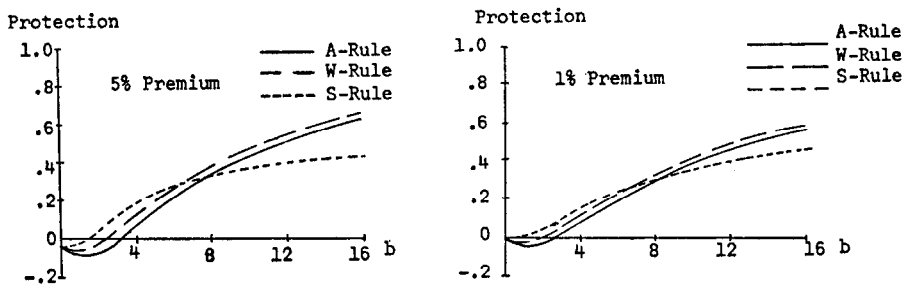

FIGURE 5—Protections corresponding to premiums of 5% and 1% when a spurious observation from $N(\mu, (1 + b)\sigma^2)$ is present in a sample of size three and σ^2 is known.

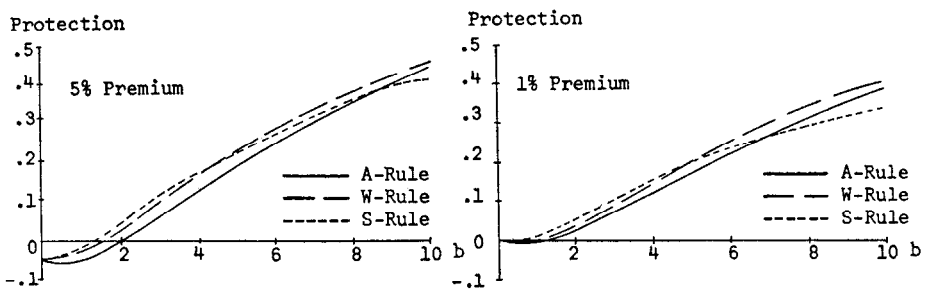

FIGURE 6—Protections corresponding to premiums of 5% and 1% when a spurious observation from $N(\mu, (1 + b)\sigma^2)$ is present in a sample of size six and σ^2 is known.

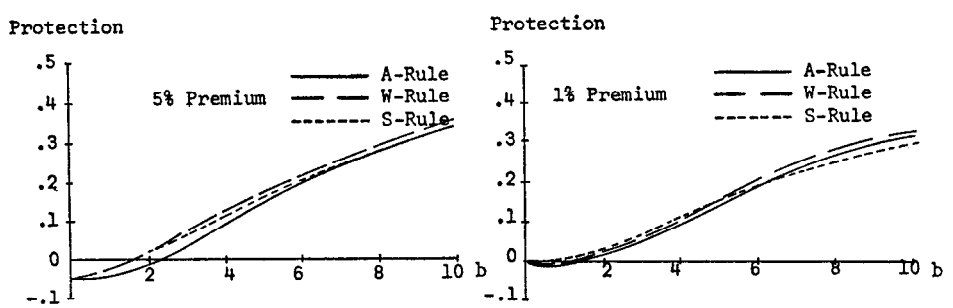

FIGURE 7—Protections corresponding to premiums of 5% and 1% when a spurious observation from $N(\mu, (1 + b)\sigma^2)$ is present in a sample of size ten and σ^2 is known.

* As one of the referees has indicated, the overall situation may be summarized to some extent by stating that the W-rule is never worst and is sometimes best, and noting that similar statements cannot be made for the A-rule or the S-rule.

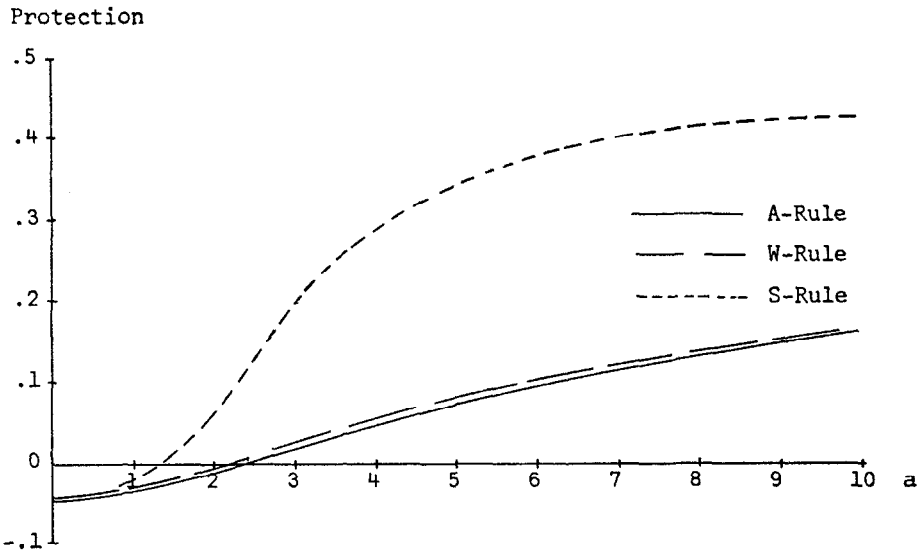

FIGURE 8—Protections corresponding to a premium 5% when a spurious observation from $N(\mu + a\sigma, \sigma^2)$ is present in a sample of size three and σ^2 is unknown.

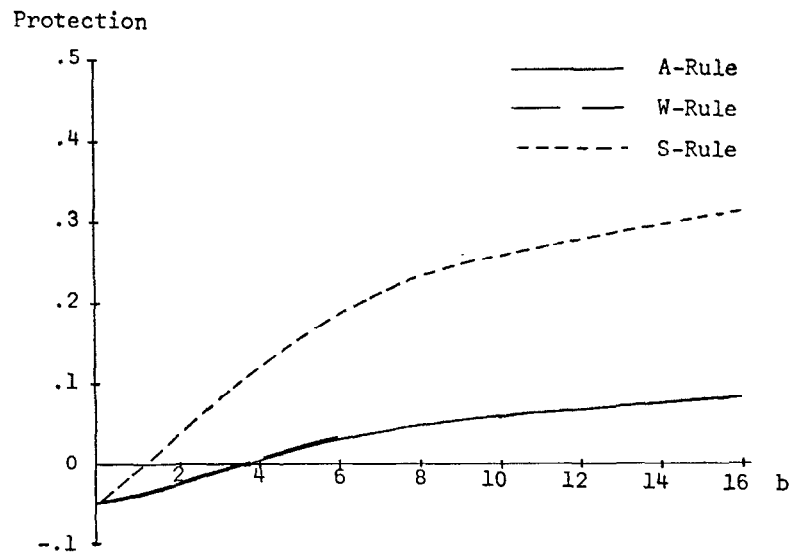

FIGURE 9—Protections corresponding to a premium of 5% when a spurious observation from $N(\mu, (1 + b)\sigma^2)$ is present in a sample of size three and σ^2 is unknown.

TABLE 1
Required constants, C, corresponding to a given premium for a sample of size three.

Premium	A-Rule	W-Rule	S-Rule
.0500	2.39038	2.30555	.98911
.0400	2.46003	2.37957	1.07104
.0300	2.54623	2.47075	1.17252
.0200	2.66184	2.59230	1.30853
.0100	2.84625	2.78487	1.52501
.0050	3.01724	2.96228	1.72493

TABLE 2
Protections corresponding to premiums of 5% and 1% when a spurious observation from $N(\mu + a\sigma, \sigma^2)$ is present in a sample of size three and σ^2 is known.

	Protections for a 5% Premium			Protections for a 1% Premium		
Bias "a"	A-Rule	W-Rule	S-Rule	A-Rule	W-Rule	S-Rule
± .5	-.065	-.063	-.054	-.015	-.014	-.011
± 1.0	-.098	-.089	-.050	-.027	-.025	-.009
± 1.5	-.126	-.104	-.016	-.048	-.035	.008
± 2.0	-.125	-.083	.051	-.051	-.036	.046
± 3.0	.033	.109	.242	.008	.049	.176
± 4.0	.364	.431	.432	.243	.298	.337
± 5.0	.667	.697	.571	.566	.603	.476
± 6.0	.831	.835	.659	.794	.804	.574
± 7.0	.900	.895	.711	.893	.889	.639
± 8.0	.930	.923	.744	.929	.923	.682
± 9.0	.946	.940	.767	.946	.940	.712
± 10.0	.956	.951	.783	.956	.951	.735

TABLE 3

Protections corresponding to premiums of 5% and 1% when a spurious observation from $N(\mu, (1 + b)\sigma^2)$ is present in a sample of size three and σ^2 is known.

	Protections for a 5% Premium			Protections for a 1% Premium		
Bias "b"	A-Rule	W-Rule	S-Rule	A-Rule	W-Rule	S-Rule
1	-.084	-.071	-.022	-.025	-.020	.006
2	-.055	-.027	.044	-.011	.002	.050
3	.008	.045	.114	.030	.050	.100
4	.081	.121	.176	.085	.109	.150
5	.154	.194	.230	.145	.171	.195
6	.221	.261	.274	.203	.230	.236
7	.282	.320	.310	.258	.284	.272
8	.336	.372	.338	.308	.334	.305
9	.384	.418	.360	.354	.379	.334
10	.426	.458	.376	.396	.419	.360
12	.498	.525	.395	.468	.488	.404
14	.556	.580	.401	.526	.545	.439
16	.603	.624	.420	.575	.591	.466

TABLE 4

A comparison of protections given by analytic and Monte Carlo methods, biased mean case, 1% premium, $n = 3$, σ^2 known. (Estimated standard errors are indicated in parentheses.)

	A - Rule		W - Rule		S - Rule	
Bias "a"	Analytic	Monte Carlo	Analytic	Monte Carlo	Analytic	Monte Carlo
± 2	-.051	-.050(.002)	-.036	-.038(.002)	.046	.049(.010)
± 4	.243	.236(.008)	.298	.298(.006)	.337	.335(.008)
± 6	.794	.798(.004)	.804	.810(.004)	.574	.576(.006)
± 8	.929	.930(.003)	.923	.927(.004)	.682	.679(.004)
± 10	.956	.952(.003)	.950	.952(.003)	.735	.719(.009)

TABLE 5

A comparison of protections given by analytic and Monte Carlo methods, biased variance case, 1% premium, $n = 3$, σ^2 known. (Estimated standard errors are indicated in parentheses.)

	A - Rule		W - Rule		S - Rule	
Bias "b"	Analytic	Monte Carlo	Analytic	Monte Carlo	Analytic	Monte Carlo
2	-.011	-.012(.001)	.002	.002(.002)	.050	.051(.004)
4	.085	.076(.004)	.109	.110(.007)	.150	.161(.008)
6	.203	.204(.004)	.230	.230(.004)	.236	.246(.008)
8	.308	.311(.006)	.334	.334(.006)	.305	.309(.010)
10	.396	.402(.008)	.419	.425(.009)	.360	.340(.009)

TABLE 6

Required constants, C, determined by Monte Carlo for premiums of 5% and 1%.

	A - Rule		W - Rule		S - Rule	
n	5% Premium	1% Premium	5% Premium	1% Premium	5% Premium	1% Premium
3	2.39	2.85	2.30	2.78	.99	1.52
4	2.51	2.99	2.34	2.87	1.09	1.62
6	2.61	3.13	2.28	2.91	1.21	1.79
8	2.66	3.19	2.18	2.90	1.25	1.90
10	2.68	3.23	2.03	2.88	1.27	1.94

TABLE 7

Protections (computed by Monte Carlo) when a spurious observation from $N(\mu + a\sigma, \sigma^2)$ is present in a sample of size n and σ^2 is known. (Estimated standard errors are indicated in parentheses.)

	5% Premium			1% Premium		
Bias "a"	A-Rule	W-Rule	S-Rule	A-Rule	W-Rule	S-Rule
			n = 4			
±1	-.080(.004)	-.073(.002)	-.039(.009)	-.023(.001)	-.020(.001)	-.006(.003)
±2	-.080(.005)	-.020(.004)	.072(.009)	-.035(.002)	-.012(.001)	.063(.004)
±3	.129(.008)	.213(.006)	.276(.010)	.064(.002)	.125(.002)	.227(.006)
±4	.465(.010)	.523(.006)	.493(.008)	.331(.006)	.403(.002)	.387(.007)
±6	.847(.004)	.833(.004)	.701(.004)	.832(.004)	.817(.004)	.634(.005)
±8	.922(.003)	.910(.005)	.796(.003)	.917(.003)	.902(.005)	.738(.004)
±10	.946(.007)	.932(.004)	.831(.003)	.946(.002)	.935(.004)	.793(.004)
			n = 6			
±1	-.072(.003)	-.056(.003)	-.033(.007)	-.018(.001)	-.016(.001)	-.007(.002)
±2	-.053(.008)	.016(.005)	.070(.008)	-.022(.001)	.009(.001)	.060(.003)
±3	.164(.006)	.271(.007)	.293(.009)	.098(.002)	.168(.002)	.219(.006)
±4	.520(.003)	.530(.005)	.489(.007)	.387(.004)	.444(.002)	.407(.007)
±6	.826(.004)	.788(.004)	.719(.004)	.823(.002)	.785(.004)	.647(.005)
±8	.897(.002)	.873(.005)	.810(.003)	.890(.004)	.873(.005)	.767(.003)
±10	.934(.002)	.905(.006)	.859(.003)	.930(.002)	.916(.004)	.820(.003)
			n = 8			
±1	-.063(.004)	-.049(.004)	-.044(.008)	-.017(.001)	-.013(.001)	-.005(.002)
±2	-.048(.005)	.048(.006)	.072(.007)	-.015(.002)	.016(.002)	.047(.003)
±3	.175(.005)	.262(.007)	.272(.008)	.099(.002)	.176(.002)	.193(.005)
±4	.493(.008)	.492(.007)	.469(.006)	.385(.004)	.434(.003)	.382(.006)
±6	.794(.003)	.746(.004)	.707(.004)	.789(.003)	.740(.004)	.639(.004)
±8	.875(.002)	.840(.007)	.810(.002)	.876(.003)	.847(.006)	.762(.003)
±10	.915(.002)	.897(.005)	.857(.006)	.915(.002)	.892(.004)	.825(.003)
			n = 10			
±1	-.061(.004)	-.040(.004)	-.038(.007)	-.015(.001)	-.012(.001)	-.005(.002)
±2	-.041(.003)	.037(.006)	.048(.007)	-.014(.002)	.018(.003)	.042(.007)
±3	.172(.007)	.236(.008)	.249(.007)	.098(.002)	.169(.003)	.191(.005)
±4	.461(.003)	.452(.006)	.446(.006)	.371(.005)	.407(.007)	.358(.006)
±6	.756(.001)	.711(.008)	.686(.003)	.757(.003)	.701(.010)	.624(.004)
±8	.849(.001)	.810(.007)	.794(.002)	.857(.001)	.825(.004)	.753(.003)
±10	.898(.001)	.876(.004)	.852(.005)	.897(.002)	.877(.004)	.822(.006)

TABLE 8

Protections (computed by Monte Carlo) when a spurious observation from $N(\mu, (1 + b)\sigma^2)$ is present in a sample of size n and σ^2 is known. (Estimated standard errors are indicated in parentheses.)

Bias "b"	5% Premium A-Rule	5% Premium W-Rule	5% Premium S-Rule	1% Premium A-Rule	1% Premium W-Rule	1% Premium S-Rule
			n = 4			
1	-.067(.004)	-.044(.004)	-.010(.005)	-.018(.001)	-.012(.001)	.008(.002)
2	-.016(.004)	.001(.006)	.047(.008)	.003(.001)	.021(.002)	.051(.004)
4	.109(.009)	.166(.004)	.200(.009)	.109(.006)	.149(.009)	.178(.009)
6	.248(.006)	.298(.007)	.299(.009)	.230(.004)	.262(.004)	.267(.008)
8	.371(.009)	.390(.009)	.375(.009)	.330(.006)	.364(.007)	.333(.009)
10	.451(.008)	.477(.008)	.447(.009)	.415(.008)	.452(.009)	.398(.010)
			n = 6			
1	-.055(.003)	-.035(.005)	-.023(.004)	-.014(.001)	-.004(.001)	.007(.002)
2	-.009(.005)	.027(.008)	.044(.007)	.011(.001)	.029(.002)	.043(.004)
4	.112(.003)	.162(.005)	.164(.008)	.105(.005)	.141(.008)	.146(.007)
6	.245(.006)	.279(.008)	.268(.009)	.233(.004)	.251(.005)	.239(.008)
8	.346(.008)	.356(.010)	.357(.009)	.310(.006)	.341(.007)	.294(.009)
10	.432(.008)	.439(.009)	.412(.009)	.395(.008)	.410(.009)	.389(.010)
			n = 8			
1	-.056(.003)	-.033(.006)	-.023(.004)	-.012(.001)	-.005(.001)	.007(.001)
2	-.010(.005)	.022(.008)	.028(.006)	.009(.001)	.027(.002)	.040(.003)
4	.102(.010)	.139(.005)	.138(.007)	.107(.006)	.125(.009)	.124(.006)
6	.213(.005)	.241(.008)	.245(.008)	.198(.003)	.226(.005)	.213(.009)
8	.314(.009)	.298(.010)	.329(.008)	.293(.005)	.309(.006)	.280(.009)
10	.372(.010)	.390(.009)	.383(.009)	.368(.008)	.379(.009)	.338(.008)
			n = 10			
1	-.048(.003)	-.034(.006)	-.028(.003)	-.012(.001)	-.003(.001)	.002(.004)
2	-.019(.002)	.028(.008)	.026(.005)	.010(.001)	.027(.002)	.029(.007)
4	.086(.003)	.112(.006)	.109(.009)	.087(.005)	.103(.008)	.103(.005)
6	.193(.005)	.194(.008)	.199(.007)	.177(.003)	.197(.004)	.195(.009)
8	.282(.007)	.287(.008)	.286(.009)	.264(.005)	.280(.007)	.241(.010)
10	.349(.009)	.351(.010)	.341(.010)	.333(.007)	.335(.008)	.306(.009)

given the experimenter an overall guide to show him exactly how each rule functions, and having this guide he may decide which rule has the characteristics most suited to his needs.

(2) If we admit the possibility of prior knowledge which can be used to put even rough bounds on the bias of a spurious observation which might occur, the experimenter would, of course, be able to choose the rule which functions best in this restricted region.

Leaving aside both the multivariate and designed experiments situations, where the field of outliers is verdant, we find no lack of open problems in the univariate case. Immediately, the problems of larger sample sizes(*), the possibility of more than one spurious observation, and non-normality arise.

However, even the somewhat restricted framework in which we have been working is not void of problems. One such problem which would be quite interesting to attack is the performance of a composite A-W-S-rule whereby the mean μ would be estimated by $\bar{y}$ if $|z_M| < C_1\sigma$, by a Semiwinsor estimator if $C_1\sigma \leq |z_M| < C_2\sigma$, by a Winsor estimator if $C_2\sigma \leq |z_M| < C_3\sigma$, and by an Anscombe estimator if $C_3\sigma \leq |z_M|$, where $C_1 \leq C_2 \leq C_3$ and $z_M = \max(-z_{(1)}, z_{(n)})$. Of course, an obvious starting question is how to optimally determine C_1, C_2, and C_3 for a given premium, if this is possible.

In the case where σ^2 is unknown, the protection afforded by the W-Rule and the A-Rule is extremely small for even moderate biases. Thus, we must echo Anscombe (1960) and say that these two rejection rules are "utterly useless and absurd" for a sample size of three. We must, however, add a note of guarded optimism in regard to the S-Rule, which performs extremely well with respect to the A and W-Rules. Of course, whether this performance will be of the same caliber for larger sample sizes is an open question.

ACKNOWLEDGEMENTS

The authors would like to express their thanks to Professor F. J. Anscombe, Yale University for many helpful comments. This research was supported by the National Science Foundation and the Wisconsin Alumni Research Foundation.

REFERENCES

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

ANSCOMBE, F. J., and BARRON, B. A. (1960). Treatment of outliers in samples of size three, *J. Res. NBS*, 70B, 141.

BOX, G. E. P., and MULLER, M. E. (1958). A note on the generation of random normal deviates, *Annals of Mathematical Statistics, 29*, 610.

DIXON, W. J. (1953). Processing data for outliers, *Biometrics, 9*, 74.

DIXON, W. J. (1960). Simplified estimation from censored normal samples, *Annals of Mathematical Statistics, 31*, 385.

GEBHARDT, F. (1964). On the risk of some strategies for outlying observations, *Annals of Mathematical Statistics, 35*, 1524.

GEBHARDT, F. (1966). On the effect of stragglers on the risk of some mean estimators in small samples, *Annals of Mathematical Statistics, 37*, 441.

GRUBBS, F. E. (1950). Sample criteria for testing outlying observations, *Annals of Mathematical Statistics, 21*, 27.

MCKAY, A. T. (1935). The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe, *Biometrika, 27*, 466.

NAIR, K. R. (1948). The distribution of the extreme deviate from the sample mean and its studentized form, *Biometrika, 35*, 118.

RIDER, P. R. (1933). Criteria for rejection of observations, *Washington University Studies (New Series, Science and Technology). No. 8*, St. Louis.

SMITH, D. E. (1966). Investigation of rejection rules for outliers in small samples from the normal distribution. Ph.D. thesis, University of Wisconsin, Madison, Wisconsin.

TIAO, G. C. and IRWIN GUTTMAN (1967). Analysis of Outliers with Adjusted Residuals, *Technometrics*, 9.

VEALE, J. R., and HUNTSBERGER, D. V. (1965). A weighted estimator for a mean when one observation may be spurious (unpublished).

(*) The problem of outliers in the setting discussed here when sample sizes are moderate or large is discussed in Tiao and Guttman (1967).