AD-A040 327 NORTH CAROLINA UNIV AT CHAPEL HILL DEPT OF STATISTICS F/G 12/1

TESTING A NORMAL SAMPLE FOR MULTIPLE OUTLIERS.(U)

NOV 76 D M HAWKINS DAHC04-74-C-0030

UNCLASSIFIED MIMEO SER-1094 ARO-11959.14.M NL

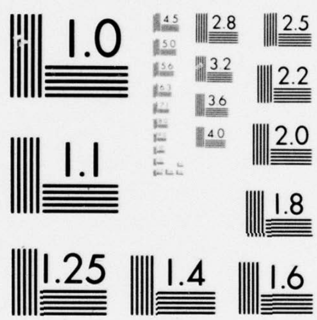
1 OF 1

AD A040327

END

DATE FILMED

7-77



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ARO-11959.4-M

(12)

TESTING A NORMAL SAMPLE FOR MULTIPLE OUTLIERS

TECHNICAL REPORT

DOUGLAS M. HAWKINS

NOVEMBER, 1976

U. S. ARMY RESEARCH OFFICE
RESEARCH TRIANGLE PARK

DAHC04-74-C-0050

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
CHAPEL HILL, NORTH CAROLINA 27514

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(14) Mimeo Ser-1094

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) Testing A Normal Sample For Multiple Outliers.		5. TYPE OF REPORT & PERIOD COVERED (9) TECHNICAL rept.
		6. PERFORMING ORG. REPORT NUMBER Mimeo Series No. 1094
7. AUTHOR(s) (10) Douglas M. Hawkins		8. CONTRACT OR GRANT NUMBER(s) (15) DAHC04-74-C-0030
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, University of North Carolina, Chapel Hill, North Carolina 27514		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Army Research Office, P. O. Box 1221, Research Triangle Park, NC 27709		12. REPORT DATE (11) November, 1976 (12) 23p.
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release: Distribution Unlimited

(18) ARO

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(19) 11959.14.M

ACCESSION for
NTIS White Section
DDC Buff Section
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. AVAIL. and/or SPECIAL
A

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Outlier, Slippage, Power, Linear Model, Regression

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Test statistics G and M for two outliers due to Murphy and Grubbs are compared with a test statistic R based on applying a single-outlier test to the sample with the most extreme observation removed. It is shown that G is generally most effective. The G and R procedures are shown to be equivalent when the number of outliers is unknown, and to give satisfactory performance for one or two outliers. An extension to the linear hypothesis is outlined.

182850

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Testing a Normal Sample for Multiple Outliers*

Douglas M. Hawkins

Department of Applied Mathematics
University of the Witwatersrand
Johannesburg, South Africa

and

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina 27514

SUMMARY

Test statistics G and M for two outliers due to Murphy and Grubbs are compared with a test statistic R based on applying a single-outlier test to the sample with the most extreme observation removed. It is shown that G is generally most effective. The G and R procedures are shown to be equivalent when the number of outliers is unknown, and to give satisfactory performance for one or two outliers. An extension to the linear hypothesis is outlined.

Key Words: Outlier, slippage, power, linear model, regression.

* Research partially supported by the U. S. Army Research Office under Grant No. DAHCO4-74-C-0030.

Introduction

The problem of identifying outliers in a sample is a very old one, but is nevertheless not yet solved entirely. A considerable amount of work has been devoted to the situation in which the underlying distribution is normal, and at most one outlier is believed to be present. It is known (Paulsen 1952, Kudo 1956) that if this outlier is known a priori to be on the right, then the optimal test statistic is the largest studentized deviation. An early paper by Pearson and Chandrasekar (1936) highlighted a severe problem associated with the use of this statistic -- the "masking effect." This effect occurs when in fact more than one outlier is present. The additional outliers mask each others' presence by so inflating the sample variance as to reduce the power of the test severely -- in some cases driving it to zero as the outliers move further to the right.

This masking effect has led to the development of several outlier test statistics aimed at testing specifically for the presence of $K>1$ outliers in the sample (Grubbs 1950, Murphy 1951, Fergusen 1961, Dixon 1950). These K-outlier statistics are competitors to the recursive procedure: Compute the single-outlier statistic for the full sample, and for each of the subsamples obtained by deleting the $1,2,3,\ldots,K$ most extreme observations, and decide on the basis of some rule using this sequence whether outliers are present, and how many and which observations they are.

The object of this paper is to propose a modification to this recursive procedure, and to compare its performance with that of the Grubbs and Murphy statistics.

Notation

Let $X_1, X_2, \ldots, X_n$ be n observations believed to come from a normal distribution $N(\xi, \sigma^2)$, ξ, σ unknown, and let $Y_1 \geq Y_2 \geq \ldots \geq Y_n$ be the corresponding order statistics. We will also suppose that there is at hand some additional information on σ^2 in the form of an independent $\sigma^2\chi_\nu^2$ variate W. If there is no such W, then set $W = \nu = 0$.

Define $$\bar{Y} = \sum_{i=1}^{n} Y_i/n$$

$$\bar{Y}_1 = \sum_{i=2}^{n} Y_i/(n-1)$$

$$\bar{Y}_{12} = \sum_{i=3}^{n} Y_i/(n-2)$$

$$S = W + \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$S_1 = W + \sum_{i=2}^{n} (Y_i - \bar{Y}_1)^2$$

$$S_{12} = W + \sum_{i=3}^{n} (Y_i - \bar{Y}_{12})^2 .$$

Initially make the somewhat artificial assumption that it is known a priori that, if outliers are present, there are exactly two, one a $N(\xi + \lambda\sigma, \sigma^2)$ and the other a $N(\xi + \delta\sigma, \sigma^2)$ variable with $\lambda, \delta > 0$. The two-outlier test statistics to be considered are $G = S_{12}/S$, $M = (Y_1 + Y_2 - 2\bar{Y})/\sqrt{S}$ and $R = (Y_2 - \bar{Y}_1)\sqrt{S_1}$.

The statistic G is a slight generalization of one proposed by Grubbs (1950) (in the original definition, $W = \nu = 0$). It seems to have no proven optimality properties, but corresponds to the F test of the Anova model for comparing the three "samples" Y_1 , Y_2 and $\{Y_3,\ldots,Y_n\}$. For this reason, it seems likely to be effective when it is not known a priori that $\lambda = \delta$.

Murphy (1951) defined the statistic M for the case $W = \nu = 0$, and proved its optimality when it is known a priori that $\lambda = \delta$. We note that this statistic is equivalent to the Anova test for comparing the two "samples" $\{Y_1,Y_2\}$ and $\{Y_3,\ldots,Y_n\}$, which in turn is equivalent to the two-sample t statistic. Of course neither G nor M has a distribution which can be derived from the conventional Anova distributions, since the "samples" are obtained by ranking the X_i .

Both statistics are extensions of the single-outlier statistic $B = (Y_1 - \bar{Y})/\sqrt{S}$ to the two-outlier situation. For the single-outlier case, Grubbs proposed the statistic $S_1/S \equiv \{1 - nB^2/(n-1)\}$.

The recursive statistic R is clearly not optimal since it does not use the sufficient statistics Y_1 , Y_2 , $\bar{Y}_{12}$ and S_{12} , but is a natural one to consider since, if Y_1 is an outlier, it corresponds to the Paulson optimal statistic for the subsample $\{Y_2,Y_3,\ldots,Y_n\}$.

The three test statistics are related, in that G and M may be expressed as functions of B and R. The relationships may be established by use of some algebraic identities from Quesenberry and David (1961) and are

$$G = \{1 - nB^2/(n-1)\}\{1 - (n-1)R^2/(n-2)\}$$

$$M = (n-2)B/(n-1) + R\{1 - nB^2/(n-1)\}^{\frac{1}{2}} . \tag{1}$$

The null joint distribution of R and B is given in Hawkins (1973), where it is shown that if $F_{n,\nu}(\cdot)$ denotes the cumulative distribution function of B based on a sample of size n, then the probability

$$\Pr[(b < B < b+db) \cap (x \le R \le B)] = nf(b)[F_{n-1,\nu}\{t(b)\} - F_{n-1,\nu}(x)]db + o(db) \tag{2}$$

where $$t(b) = nb/[(n-1)\{1 - nb^2/(n-1)\}^{\frac{1}{2}}]$$

and $$f(b) = \frac{\Gamma\{\frac{1}{2}(n+m-1)\}}{\Gamma\{\frac{1}{2}(n+m-2)\}}\{n/(n-1)\pi\}^{\frac{1}{2}}\{1 - nb^2/(n-1)\}^{\frac{1}{2}(n+m-4)} .$$

The range is $0 \le x \le t(b) \le \{(n-1)/n\}^{\frac{1}{2}}$, the constraints $R \le B$ and $x \le t(b)$ being imposed by the condition $Y_1 \ge Y_2$.

The Grubbs test has critical region $G < G_0$, whose size is

$$\Pr[G < G_0] = \Pr[\{1 - nB^2/(n-1)\}\{1 - (n-1)R^2/(n-2)\} < G_0]$$
$$= \Pr[\{1 - nB^2/(n-1)\} < G_0\{1 - (n-1)R^2/(n-2)\}^{-1}] . \tag{3}$$

For any G_0, we may integrate (2) over the region defined by (3), and so get this probability. In this way, the distribution function of G may be found and fractiles deduced. For the case $W = \nu = 0$, fractiles were found by Grubbs, who found the density of G directly, though in terms of a multiple integral over a complicated region.

The Murphy test has critical region $M > M_0$, whose size is

$$\Pr[M > M_0] = \Pr[(n-2)B/(n-1) + R\{1 - nB^2/(n-1)\}^{\frac{1}{2}} > M_0] \tag{4}$$

and so the null distribution of M may be found in the same way.

Finally, the size of the region $R > R_0$ may be found from (2), and its fractiles deduced.

Some fractiles of G, M and R are listed in Table 1. Those fractiles listed both here and in Grubbs (1950) agree. So far as this author is aware, this is the first table of exact fractiles of M and R.

Figure 1 shows the critical regions of each of these tests for the case $n = 10$, $\nu = 0$, $\alpha = .05$. Recalling that a large value of B suggests that Y_1 is aberrant, while a large value of R indicates that Y_2 is aberrant, we can interpret these regions as follows:

i) Each test is best in some region of (B,R) in the sense of rejecting the null hypothesis, while both other tests accept it.

ii) If $Y_1 \to \infty$ while $\{Y_2,Y_3,\ldots,Y_n\}$ remain constant, then ultimately $\{B,R\}$ falls in the acceptance region of the test based on M, and the critical region of that based on G. This means that G tends to be significant when in fact only a single outlier is present. This may be a shortcoming in the use of G. The fact that M may become non-significant when Y_1 becomes more aberrant is interesting, and possibly surprising. In fact, if we let $Y_1 \to \infty$, $\{Y_2,Y_3,\ldots,Y_n\}$ remaining fixed, it can be shown quite easily that $M \to C = (n-2)/\{n(n-1)\}^{\frac{1}{2}}$. Thus if the fractile $M_0 > C$, then M will ultimately become non-significant. This means that M suffers from a type of masking effect, in that if $M_0 > C$, it is not effective for finding two outliers whose means are markedly dissimilar.

iii) By contrast, the critical region of R is specific for two outliers, and is not affected by the relative magnitude of B and R.

Performance of the tests.

In this section, measures of the performance of the three tests in locating exactly two outliers are considered. The best criterion would be the probability of a correct decision — that is the probability that the two contaminants are the first two order statistics of the sample, and that (B,R) falls in the critical region. The evaluation of this probability however seems to involve some rather difficult distributional problems, familiar from the single-outlier case (David and Paulson 1965), and so an approximation similar to that of David and Paulson is proposed.

As a preliminary, suppose that

$$T_1 \sim N(\beta,1)$$

$$T_2 \sim N(\gamma,1)$$

$$V \sim \chi^2_N$$

all three terms being mutually independent. Let

$$U = T_1/V^{\frac{1}{2}}$$

$$Z = T_2/\{V(1+U^2)\}^{\frac{1}{2}} .$$

Standard transformation methods then show that the joint density of U and Z is

$$f(u,z) = 4K \sum_{m=0}^{\infty} \frac{(\beta\sqrt{2})^m u^m}{m!(1+u^2)^{\frac{1}{2}(N+m+1)}} \sum_{i=0}^{\infty} \frac{\Gamma[\frac{1}{2}(N+i+m+2)](\gamma\sqrt{2})^i z^i}{i!(1+z^2)^{\frac{1}{2}(N+i+m+2)}}$$

where $K = \exp -\frac{1}{2}(\beta^2 + \gamma^2)/\{4\pi\Gamma(\frac{1}{2}N)\}$.

From this $\Pr[(U>b) \cap \{Z > a(U)\}]$ is

$$K \int_{v=0}^{(1+b^2)^{-1}} \sum_{m=0}^{\infty} (\beta\sqrt{2})^m v^{\frac{1}{2}(N-2)} (1-v)^{\frac{1}{2}(m-1)} \frac{\Gamma\{\frac{1}{2}(N+m+1)\}}{m!}$$
$$\sum_{i=0}^{\infty} \frac{(\gamma\sqrt{2})^i \Gamma\{\frac{1}{2}(i+1)\}}{i!} I_{\{1+a^2(v)\}^{-1}} \{\tfrac{1}{2}(N+m+1), \tfrac{1}{2}(i+1)\} dv \quad (5)$$

where $I_x(a,b)$ denotes the incomplete beta ratio, and the variable of integration $v = (1 + u^2)^{-1}$.

Returning to the problem at hand, suppose that

$$X_1 \sim N(\xi+\delta\sigma, \sigma^2)$$
$$X_2 \sim N(\xi+\lambda\sigma, \sigma^2)$$
$$X_i \sim N(\xi,\sigma^2) \qquad i = 3,\ldots,n$$
$$W \sim \chi^2_\nu .$$

Let

$$\overline{X} = \sum_{i=1}^{n} X_i/n$$
$$\overline{X}_1 = \sum_{i=2}^{n} X_i/(n-1).$$

Make the Helmert orthogonal transformation, defining

$$T_1 = \{n/(n-1)\}^{\frac{1}{2}}(X_1 - \overline{X}) \sim N(\beta\sigma,\sigma^2)$$
$$T_2 = \{(n-1)/(n-2)\}^{\frac{1}{2}}(X_2 - \overline{X}_1) \sim N(\gamma\sigma,\sigma^2)$$

where $$\beta = \{n/(n-1)\}^{\frac{1}{2}}\delta - \{n(n-1)\}^{-\frac{1}{2}}\lambda$$

$$\gamma = \{(n-2)/(n-1)\}^{\frac{1}{2}}\lambda.$$

Letting $$S = W + \sum_{i=1}^{n} (X_i - \bar{X})^2$$

we identify T_1 and T_2 with the T_1 and T_2 of the preceding theory, V with $S - T_1^2 - T_2^2$, and N with $n+\nu-2$.

Then $$U = T_2/(S - T_1^2 - T_2^2)^{\frac{1}{2}}$$

$$Z = T_1/(S - T_1^2)^{\frac{1}{2}}$$

$$T_1/S^{\frac{1}{2}} = Z/(1 + Z^2)^{\frac{1}{2}}$$

$$T_2/(S - T_1^2)^{\frac{1}{2}} = U/(1 + U^2)^{\frac{1}{2}} .$$

Now $$B' = \{(n-1)/n\}^{\frac{1}{2}}T_1/S^{\frac{1}{2}}$$

$$R' = \{(n-2)/(n-1)\}^{\frac{1}{2}}T_2/(S - T_1^2)^{\frac{1}{2}}$$

are simply the B and R statistics obtained by substituting X_1 for Y_1 and X_2 for Y_2 .

Let C be the critical region for any particular test, and let

$$H(\delta,\lambda) = \Pr[(B',R') \in C].$$

This differs from $\Pr[\{(B',R') \in C\} \cap \{X_1=Y_1, X_2=Y_2\}]$ by the factor

$$\Pr[(X_1=Y_1) \cap (X_2=Y_2) | (B',R') \in C]. \tag{6}$$

Now we may argue, as in Anscombe (1960) that if $\delta > \lambda >> 0$, this conditional probability is close to 1. Alternatively, we may note as in David and Paulson that if $B' \geq [(n-2)/2n]^{\frac{1}{2}}$, $R' \geq [(n-3)/(2n-2)]^{\frac{1}{2}}$,

$\Pr[(X_1=Y_1) \cap (X_2=Y_2) | (B',R')] = 1.$

Either reasoning shows that the probability of a correct decision:

$$\Pr[(X_1=Y_1) \cap (X_2=Y_2) \cap (B',R') \in C] + \Pr[(X_1=Y_2) \cap (X_2=Y_1) \cap (B',R') \in C]$$

is well-approximated by $H(\delta,\lambda) + H(\lambda,\delta)$.

Table 2 shows the values of $H(\delta,\lambda) + H(\lambda,\delta)$ for the case $n = 10$, $\nu = 0$, $\alpha = .05$ whose critical regions were sketched in Figure 1. These values confirm the impressions given by Figure 1. If $\delta=\lambda$, H is the best test statistic, but its power deteriorates if δ is markedly different from λ. Of interest is the good performance of G; however its power when $\lambda=0$ is a warning of the possibility of rejecting H_0 when in fact only one outlier is present. The recursive test is never best, but is never poor when, in fact, two outliers are present.

Extension to an unknown number of outliers

Suppose now that it is not known a priori whether the sample contains 0, 1 or 2 outliers. A question then arises of how to adapt the test procedures.

The Grubbs statistic is equivalent to an Anova for the three groups Y_1, Y_2 and $\{Y_3,\ldots,Y_n\}$. It might be adapted by setting up the Anova for comparing Y_1 with $\{Y_2,\ldots,Y_n\}$ whose test statistic is equivalent to $S_1/S = \{1 - nB^2/(n-1)\}$. Then S_1 is further partitioned for testing Y_2 against $\{Y_3,\ldots,Y_n\}$ yielding a test statistic equivalent to $S_{12}/S_1 = \{1 - (n-1)R^2/(n-2)\}$. The decision as to the number of outliers would then be equivalent to a sequential test based on B and R.

The recursive test which has received most attention (McMillan 1971, Hawkins 1973) is:

i) Compute B. If this is non-significant, conclude that there are no outliers.

ii) Otherwise compute R, and depending on its significance, conclude that there are one, or two outliers.

It is now well-known that this procedure is defective because of the masking effect. However this type of problem is very familiar in other contexts – for example deciding by what order polynomial to detrend a time series (Anderson 1971), and may be resolved very simply by reversing the order of testing. This yields the procedure:

i) Compute R. If it is significant conclude that there are two outliers.

ii) Otherwise, compute B, and depending on its significance, conclude that there are one, or no outliers.

The fractiles needed for this procedure may be computed by integrating (2) over the two regions: $C_2 : R>R_0$ and $C_1 : R<R_0, B>B_0$. The test procedure is not fully defined until one makes a decision as to how to partition α, the overall probability of type I error, between these two regions. One reasonable partition is to let each have size $\alpha/2$.

A sample calculation for the case $n = 10$, $\nu = 0$, $\alpha = .05$ yields the fractiles $R_0 = 0.7484$; $B_0 = 0.7596$. Let $H_i(\delta,\lambda)$ denote $\Pr[(B',R') \in C_i]$ $i = 1,2$ when in fact outliers are present, then $H_i(\delta,\lambda) + H_i(\lambda,\delta)$ approximates the probability of identifying two outliers $(i=2)$ or one outlier $(i=1)$.

Table 3 lists these approximations for this sample case, and shows that the performance of the recursive procedure is indeed satisfactory.

The recursive procedure extends in a natural way to test for the possibility of $1,2,\ldots,K$ outliers. The sequence of recursive statistics omitting the most aberrant $0,1,2,\ldots,K-1$ observations is computed, and tested for significance in reverse order. No fractiles for the case $k>2$ are as yet available, though it is likely that the approach used in Hawkins (1973) could be extended. It will be shown below that conservative fractiles may be obtained by the use of the Bonferroni inequality.

The Murphy statistic extends in an obvious way to accommodate an unknown number of outliers: For $k = 1,2,\ldots,K$, find the two-sample t statistic for comparing $\{Y_1,\ldots,Y_k\}$ with $\{Y_{k+1},\ldots,Y_n\}$. If the largest of these, corresponding to $k = k_0$, say, is significant, then conclude that k_0 outliers are present. This statistic, in the case $k=n$, corresponds to a test statistic used in Automatic Interaction Detection (see for example Kass 1975, Scott and Knott 1976).

However since Table 2 suggests that M is never markedly superior to G, and may in fact be markedly inferior, no attempt has been made to study the performance of this procedure. Note that for $K=2$ its fractiles could be deduced from the basic joint distribution (2).

Extension to the linear model

Suppose now that under H_0 the X_i are generated by a linear model

$$X_i \sim N[\underset{\sim}{w}_i \underset{\sim}{\beta},\ \sigma^2] \tag{7}$$

where $\underset{\sim}{\beta}$ is a $p\times1$ vector of unknown regression coefficient and σ^2 is unknown. The alternative model that outliers are present may be formulated as

$$X_i \sim N[\underset{\sim}{w}_i \underset{\sim}{\beta}+\lambda_i,\ \sigma^2] \tag{8}$$

where $\lambda_i = 0$ for the good observations. The sign of λ_i may, or may not be fixed a priori, depending on circumstances.

The model 7 may be rewritten in terms of an indicator vector as follows: Let $\underset{\sim}{w}_i^* = (\underset{\sim}{w}_i: 0,0\ldots0,1,0\ldots0)$, the 1 occuring in position i of the indicator vector, and let

$$\underset{\sim}{\beta}^{*\prime} = (\beta',\lambda_1,\lambda_2,\ldots,\lambda_n)$$

Then
$$X_i \sim N[\underset{\sim}{w}_i^* \underset{\sim}{\beta}^*,\ \sigma^2]. \tag{9}$$

On the face of it, there seem to be two possible approaches to the problem of handling outliers here. One proceeds by eliminating outliers: any significant outlier is deleted from the sample, and the usual estimates of $\underset{\sim}{\beta}$ and σ^2 in 7 computed from the remaining data. The other method is to use the model (9), estimating $\underset{\sim}{\beta}$, σ^2 and those λ_i that tests indicate to be non-zero.

In fact, the two approaches are identical – the matrix operations involved in deleting X_i from the sample and updating the regression (Beckman and Trussell 1975) are identical to those involved in introducing λ_i to the regression.

Given this equivalence of the two methods, it is convenient for present purposes to work in terms of the second formulation.

In matrix form the model is

$$\underset{\sim}{X} \sim N[(\underset{\sim}{W}:\underset{\sim}{I}_n)\begin{pmatrix}\beta\\ \lambda\end{pmatrix},\ \sigma^2 \underset{\sim}{I}_n].$$

In solving to find regression coefficients, form

$$(\underset{\sim}{W}:\ \underset{\sim}{X}:\ \underset{\sim}{I}_n)'(\underset{\sim}{W}:\ \underset{\sim}{X}:\ \underset{\sim}{I}_n) = \begin{bmatrix} \underset{\sim}{W}'\underset{\sim}{W} & \underset{\sim}{W}'\underset{\sim}{X} & \underset{\sim}{W}' \\ \underset{\sim}{X}'\underset{\sim}{W} & \underset{\sim}{X}'\underset{\sim}{X} & \underset{\sim}{X}' \\ \underset{\sim}{W} & \underset{\sim}{X} & \underset{\sim}{I} \end{bmatrix}.$$

Introducing $\underset{\sim}{W}$ into the regression by sweeping on the first p rows and columns transforms this matrix to

$$\begin{bmatrix} -(\underset{\sim}{W}'\underset{\sim}{W})^{-1} & (\underset{\sim}{W}'\underset{\sim}{W})^{-1}\underset{\sim}{W}'\underset{\sim}{X} & (\underset{\sim}{W}'\underset{\sim}{W})^{-1}\underset{\sim}{W}' \\ \underset{\sim}{X}'\underset{\sim}{W}(\underset{\sim}{W}'\underset{\sim}{W})^{-1} & \underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{X} & \underset{\sim}{X}'\underset{\sim}{A} \\ \underset{\sim}{W}(\underset{\sim}{W}'\underset{\sim}{W})^{-1} & \underset{\sim}{A}\underset{\sim}{X} & \underset{\sim}{A} \end{bmatrix}$$

where $\underset{\sim}{A} = \underset{\sim}{I}_n - \underset{\sim}{W}(\underset{\sim}{W}'\underset{\sim}{W})^{-1}\underset{\sim}{W}'$, so that $S^2 = \underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{X}$ is the residual sum of squares of the regression of $\underset{\sim}{X}$ on $\underset{\sim}{W}$ and $\underset{\sim}{e} = \underset{\sim}{A}\underset{\sim}{X}$ is the vector of residuals.

Let a_{ij} denote the i,jth element of $\underset{\sim}{A}$, and e_i the ith element of $\underset{\sim}{e}$. Then questions about the introduction of one or more λ_i into the regression focus on the matrix

$$\begin{pmatrix} S^2 & \underset{\sim}{e}' \\ \underset{\sim}{e} & \underset{\sim}{A} \end{pmatrix}.$$

The single-outlier test would be based on the maximum of r_{X_i} — the partial correlation between X and indicator variable i. This is

$$r_{X_i} = e_i/(a_{ii}S)^{\frac{1}{2}}$$

a well-known outlier statistic (Ellenberg 1973). In the case $w_i = 1$,

$p=1$, $r_{\lambda_i} = \{n/(n-1)\}^{\frac{1}{2}}B$, so this statistic generalizes B.

Unlike the case $p=1$, the case $p>1$ poses difficulties for testing for the presence of two or more outliers: — the very difficulties that beset the regression subset problem, namely that the best subset of size k to include in the regression does not necessarily contain the best subset of size $k-1$.

If the number of outliers k is specified a priori, then no conceptual difficulty arises. The best subset of k predictors amongst the λ_i is found by using a branch and bound multiple regression algorithm. The analysis of variance then yields a test statistic equivalent to G.

The use of the recursive procedure is equivalent to a stepwise regression in which one introduces K of the λ_i in order of significance, and then proceeds to eliminate them. As in stepwise regression, a conservative test may be set up by concluding that there are fewer than k outliers if the partial correlation corresponding to the kth is not significant at the $\alpha/\{K(n-k+1)\}$ level. Experience with multiple regression shows that there is no guarantee that this procedure will identify the correct outliers, especially since A has rank only $n-p$, and so is highly multicollinear.

For the case $W_i \equiv 1$, $p=1$, the Bonferroni bound is conservative — for example in the test case considered earlier, the .05 fractile of R corresponds to a Bonferroni fractile with $\alpha \doteq .1$. Considerable experience is still needed to establish whether this procedure is effective.

REFERENCES

ANDERSON, T. W., (1971), *The Statistical Analysis of Time Series*. New York: John Wiley and Sons.

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

BECKMAN, R. J. and TRUSSELL, H. J., (1975). The distribution of an arbitrary studentized residual, and the effects of updating in multiple regression. *Journal of the American Statistical Association, 69*, 199-201.

DAVID, H. A. and PAULSON, A. S., (1965). The performance of several tests for outliers. *Biometrika, 52*, 429-436.

DIXON, W. J., (1950). Analysis of extreme values. *Annals of Mathematical Statistics, 21*, 488-506.

ELLENBERG, J. H., (1973). The joint distribution of the standardized least squares residuals from a general linear regression. *Journal of the American Statistical Association, 68*, 941-943.

FERGUSON, T. S., (1961). On the rejection of outliers. *Proceedings, Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 253-287.

GRUBBS, F. E., (1950). Sample criteria for testing outlying observations. *Annals of Mathematical Statistics, 21*, 27-58.

HAWKINS, D. M., (1973). Repeated testing for outliers. *Statistica Neerlandica, 27*, 1-10.

KASS, G. V., (1975). Significance testing in Automatic Interaction Detection (A.I.D.). *Applied Statistics, 24*, 178-189.

KUDO, A., (1956). On the testing of outlying observations. *Sankhya, 17*, 67-76.

McMILLAN, R. G., (1971). Tests for one or two outliers in normal samples with unknown variance. *Technometrics, 13*, 87-100.

MURPHY, R. B., (1951). On tests for outlying observations. Ph.D. Thesis, Princeton University.

PAULSON, E., (1952). An optimum solution to the k-sample slippage problem for the normal distribution. *Annals of Mathematical Statistics, 23*, 610-616.

PEARSON, E. S. and CHANDRASEKAR, C., (1936). The efficiency of statistical tools and a criterion for the rejection of outlying observations. *Biometrika, 28*, 308-320.

QUESENBERRY, C. P. and DAVID, H. A., (1961). Some tests for outliers. *Biometrika*, *48*, 379-390.

SCOTT, A. J. and KNOTT, M., (1976). An approximate test for use with AID. *Applied Statistics*, *25*, 103-106.

TABLE 1

CRITICAL VALUES OF G, M AND R

		0				5				15				30			
		0.1	0.05	0.01	0.001	0.1	0.05	0.01	0.001	0.1	0.05	0.01	0.001	0.1	0.05	0.01	0.001
5	G	0.036	0.017	0.002	0.0	0.407	0.330	0.204	0.104	0.696	0.638	0.524	0.397	0.824	0.788	0.710	0.614
	M	1.024	1.050	1.080	1.092	0.785	0.838	0.928	1.003	0.564	0.616	0.712	0.811	0.427	0.471	0.554	0.647
	R	0.810	0.840	0.864	0.866	0.496	0.555	0.661	0.756	0.318	0.361	0.445	0.540	0.232	0.264	0.327	0.402
6	G	0.090	0.055	0.017	0.003	0.423	0.351	0.230	0.127	0.688	0.633	0.525	0.403	0.817	0.781	0.706	0.610
	M	1.044	1.077	1.120	1.144	0.827	0.879	0.966	1.044	0.610	0.661	0.756	0.854	0.468	0.512	0.595	0.688
	R	0.771	0.815	0.869	0.892	0.505	0.560	0.660	0.758	0.336	0.377	0.458	0.550	0.247	0.278	0.340	0.413
7	G	0.146	0.100	0.043	0.012	0.440	0.372	0.255	0.150	0.683	0.632	0.528	0.411	0.810	0.775	0.701	0.610
	M	1.048	1.085	1.138	1.172	0.850	0.901	0.986	1.066	0.641	0.691	0.784	0.881	0.497	0.541	0.623	0.716
	R	0.730	0.780	0.853	0.896	0.506	0.557	0.653	0.751	0.347	0.386	0.463	0.553	0.259	0.288	0.348	0.420
8	G	0.197	0.146	0.074	0.028	0.457	0.393	0.279	0.173	0.682	0.632	0.532	0.419	0.806	0.771	0.699	0.609
	M	1.044	1.083	1.143	1.187	0.863	0.912	0.998	1.078	0.662	0.711	0.802	0.898	0.518	0.562	0.644	0.735
	R	0.695	0.746	0.828	0.888	0.503	0.551	0.643	0.740	0.354	0.392	0.466	0.553	0.267	0.296	0.354	0.424
9	G	0.244	0.189	0.107	0.048	0.474	0.413	0.302	0.195	0.682	0.634	0.538	0.428	0.803	0.768	0.698	0.611
	M	1.036	1.076	1.140	1.193	0.869	0.917	1.002	1.083	0.677	0.725	0.814	0.909	0.535	0.578	0.659	0.749
	R	0.665	0.715	0.801	0.872	0.498	0.544	0.632	0.728	0.359	0.395	0.467	0.551	0.273	0.301	0.358	0.426
10	G	0.284	0.229	0.140	0.070	0.490	0.431	0.324	0.217	0.683	0.637	0.544	0.437	0.799	0.767	0.698	0.613
	M	1.027	1.066	1.134	1.192	0.871	0.918	1.002	1.084	0.688	0.735	0.822	0.916	0.548	0.590	0.670	0.759
	R	0.639	0.687	0.774	0.853	0.493	0.536	0.621	0.715	0.362	0.397	0.466	0.548	0.278	0.305	0.361	0.427
11	G	0.321	0.265	0.172	0.094	0.505	0.449	0.344	0.237	0.686	0.641	0.551	0.447	0.799	0.765	0.698	0.615
	M	1.015	1.055	1.125	1.188	0.870	0.916	1.000	1.082	0.695	0.741	0.827	0.920	0.558	0.599	0.678	0.767
	R	0.616	0.662	0.749	0.832	0.486	0.528	0.609	0.701	0.364	0.397	0.465	0.545	0.282	0.308	0.363	0.428
12	G	0.353	0.298	0.203	0.119	0.520	0.466	0.363	0.257	0.688	0.644	0.557	0.456	0.797	0.765	0.700	0.618
	M	1.002	1.043	1.114	1.181	0.867	0.913	0.995	1.080	0.700	0.746	0.830	0.922	0.566	0.607	0.685	0.772
	R	0.596	0.640	0.725	0.811	0.479	0.519	0.598	0.688	0.364	0.397	0.462	0.540	0.284	0.311	0.364	0.428
13	G	0.383	0.328	0.232	0.143	0.534	0.481	0.382	0.276	0.691	0.649	0.564	0.465	0.797	0.765	0.701	0.621
	M	0.989	1.030	1.102	1.173	0.862	0.908	0.989	1.074	0.703	0.748	0.831	0.922	0.572	0.612	0.689	0.776
	R	0.578	0.621	0.703	0.791	0.473	0.511	0.588	0.675	0.364	0.396	0.459	0.536	0.286	0.312	0.364	0.427
14	G	0.409	0.355	0.259	0.167	0.547	0.496	0.399	0.295	0.695	0.654	0.571	0.474	0.796	0.765	0.702	0.624
	M	0.976	1.018	1.091	1.163	0.857	0.902	0.983	1.068	0.705	0.749	0.831	0.921	0.577	0.617	0.693	0.778
	R	0.562	0.603	0.684	0.771	0.466	0.503	0.577	0.663	0.364	0.395	0.457	0.531	0.289	0.314	0.364	0.426
15	G	0.433	0.380	0.284	0.190	0.559	0.510	0.415	0.312	0.698	0.658	0.578	0.483	0.796	0.766	0.704	0.628
	M	0.964	1.005	1.079	1.153	0.851	0.895	0.976	1.061	0.706	0.749	0.830	0.919	0.581	0.620	0.695	0.780
	R	0.548	0.587	0.665	0.752	0.460	0.496	0.567	0.651	0.363	0.393	0.453	0.526	0.290	0.314	0.364	0.425
20	G	0.525	0.479	0.390	0.292	0.612	0.569	0.485	0.389	0.717	0.682	0.610	0.524	0.799	0.772	0.716	0.645
	M	0.904	0.944	1.019	1.099	0.818	0.860	0.937	1.021	0.700	0.740	0.816	0.901	0.590	0.627	0.698	0.778
	R	0.491	0.524	0.593	0.673	0.431	0.461	0.524	0.600	0.355	0.382	0.436	0.502	0.292	0.314	0.360	0.416
25	G	0.589	0.548	0.467	0.374	0.653	0.615	0.539	0.450	0.735	0.703	0.639	0.560	0.804	0.780	0.728	0.663
	M	0.853	0.893	0.966	1.046	0.785	0.825	0.899	0.981	0.687	0.724	0.797	0.878	0.589	0.624	0.692	0.769
	R	0.450	0.480	0.541	0.614	0.406	0.433	0.490	0.558	0.346	0.370	0.419	0.480	0.290	0.311	0.353	0.406
30	G	0.636	0.600	0.526	0.438	0.685	0.651	0.582	0.499	0.751	0.722	0.664	0.590	0.811	0.788	0.740	0.679
	M	0.810	0.848	0.920	1.000	0.755	0.793	0.864	0.944	0.671	0.707	0.776	0.854	0.585	0.618	0.682	0.757
	R	0.421	0.447	0.502	0.569	0.385	0.410	0.462	0.525	0.335	0.358	0.403	0.460	0.287	0.306	0.346	0.396

TABLE 2
PERFORMANCE OF G, M AND R

		0	1	2	3	4	5	6	7	8	9
0	G	0.001									
	M	0.002									
	R	0.001									
1	G	0.004	0.012								
	M	0.006	0.014								
	R	0.003	0.010								
2	G	0.007	0.028	0.077							
	M	0.017	0.036	0.073							
	R	0.006	0.023	0.067							
3	G	0.009	0.047	0.143	0.283						
	M	0.049	0.095	0.152	0.246						
	R	0.007	0.038	0.123	0.254						
4	G	0.010	0.060	0.198	0.407	0.594					
	M	0.110	0.206	0.293	0.393	0.523					
	R	0.008	0.046	0.163	0.363	0.552					
5	G	0.010	0.062	0.224	0.481	0.709	0.844				
	M	0.196	0.361	0.482	0.577	0.676	0.782				
	R	0.008	0.049	0.181	0.418	0.659	0.810				
6	G	0.008	0.055	0.221	0.509	0.762	0.900	0.962			
	M	0.284	0.522	0.671	0.754	0.817	0.876	0.931			
	R	0.008	0.049	0.187	0.440	0.702	0.868	0.946			
7	G	0.005	0.044	0.198	0.497	0.782	0.919	0.977	0.993		
	M	0.351	0.646	0.814	0.881	0.916	0.942	0.967	0.985		
	R	0.008	0.049	0.188	0.447	0.722	0.886	0.964	0.989		
8	G	0.004	0.032	0.164	0.457	0.770	0.931	0.978	0.996	0.998	
	M	0.389	0.718	0.897	0.951	0.970	0.979	0.987	0.994	0.997	
	R	0.008	0.049	0.188	0.448	0.730	0.903	0.965	0.993	0.998	
9	G	0.002	0.022	0.128	0.401	0.734	0.929	0.980	0.996	0.999	0.999
	M	0.405	0.751	0.935	0.981	0.990	0.994	0.996	0.998	0.998	0.998
	R	0.008	0.049	0.187	0.448	0.731	0.911	0.971	0.992	0.999	0.998

TABLE 3
PERFORMANCE OF R AND B

	0	1	2	3	4	5	6	7	8	9
R	0.001									
B	0.003									
R	0.002	0.006								
B	0.009	0.015								
R	0.004	0.014	0.042							
B	0.032	0.038	0.036							
R	0.005	0.024	0.081	0.178						
B	0.095	0.116	0.076	0.047						
R	0.005	0.029	0.111	0.269	0.440					
B	0.198	0.271	0.196	0.083	0.036					
R	0.005	0.031	0.126	0.320	0.550	0.718				
B	0.307	0.463	0.386	0.191	0.056	0.018				
R	0.005	0.031	0.131	0.341	0.600	0.795	0.900			
B	0.380	0.619	0.581	0.343	0.121	0.026	0.007			
R	0.005	0.031	0.131	0.348	0.621	0.824	0.931	0.975		
B	0.414	0.703	0.714	0.485	0.209	0.052	0.009	0.002		
R	0.005	0.031	0.131	0.349	0.628	0.842	0.937	0.983	0.995	
B	0.425	0.736	0.779	0.578	0.287	0.087	0.016	0.002	0.001	
R	0.005	0.031	0.131	0.349	0.629	0.849	0.947	0.982	0.997	0.998
B	0.427	0.744	0.801	0.621	0.337	0.118	0.025	0.004	0.001	0.000

FIGURE 1. CRITICAL REGIONS.