TESTS FOR ONE OR TWO OUTLIERS

by

Robert G. McMillan

Institute of Statistics
Mimeograph Series No. 613
August 1968

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (continued)

Page

1. INTRODUCTION AND LITERATURE SURVEY

1.1 The Outlier Problem

The problem of treating outlying observations has been approached in various contexts and with various objectives. Usually it is assumed that a sample has been observed in which one or more of the observations may be from a distribution other than the distribution of chief interest. In chemical laboratories, for example, it is common practice to monitor the precision of an analytical method with a series of analyses on a homogeneous batch of material. These analyses may be used to provide an estimate of the true batch concentration, and this estimate may also be used in the calibration of other methods. Occasionally, because of inexperienced personnel or other causes, grossly deviant analyses occur. Such results are frequently helpful in drawing attention to conditions needing attention. Beyond this they are likely to be harmful in analysis of the data, and their rejection, or at least isolation, is desirable.

For normally distributed observations, the outliers are usually assumed to differ in mean (location contamination) or in variance (scalar contamination) from the other observations. Some studies have considered the effect of rejection of observations on the estimation of parameters. In other studies the detection of outliers has been an objective in itself.

1.2 Review of Literature

Since reviews of the literature on this subject are available elsewhere, e.g. Anscombe (1960) and Ferguson (1961), only the contributions most relevant to the present study will be mentioned here. Residuals have long been used in outlier detection. Anscombe states that Wright (1884) suggests that an observation whose residual exceeds in magnitude five times the probable error should be rejected. This rule is studentized by Thompson (1935), who proposes that in a normal sample, $x_1, x_2, \ldots, x_n$, those observations for which $|x_i - \bar{x}|/s$ exceeds some constant be classed as outliers, where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, and $s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$. Thompson provides a table of constants which allows the average number of rejections per sample free of outliers to be predetermined. Pearson and Chandra Sekar (1936) point out that $(x_i - \bar{x})/s$ is likely to be inefficient for detection of outliers if there is more than one outlier in the sample. The presence of additional outliers tends to increase s and mask all of them. They show that if $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ are the ordered observations, there are algebraic maxima which $(x_{(n)} - \bar{x})/s$, $(x_{(n-1)} - \bar{x})/s$, etc. cannot exceed.

The distribution of $(x_{(n)} - \bar{x})/\sigma$ is obtained by McKay (1935). Simpler derivations are given by Nair (1948) and Grubbs (1950). Nair tabulates the probatility integral of this statistic for $n \leq 9$. Grubbs furnishes the upper .5%, 1%, 5%, and 10% values, $n \leq 25$, for use as rejected constants. For use when the variance is unknown, Nair determines upper 1% and 5% points for $(x_{(n)} - \bar{x})/s_\nu$, where s_ν^2 is a mean square estimate of variance with ν d.f. and independent of $x_1, x_2, \ldots, x_n$. Grubbs obtains the distribution of a statistic equivalent to $(x_{(n)} - \bar{x})/s$ and upper 1%, 2.5%, 5%, and 10% values. Quesenberry and David (1961) give constants for rejection of outliers based on $(x_{(n)} - \bar{x}) / \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 + \nu s_\nu^2}$, thus using all the data to estimate σ when external information is available.

Procedures in which $x_{(n)}$ is classed as an outlier if $(x_{(n)}-\bar{x})/s$ exceeds some constant are shown by Murphy (1951), Paulson (1952), and Kudō (1956) to have a certain optimum property. The precise nature of the optimality will be explained in Chapter 3. All three of the developments are based on restriction to procedures which are invariant under change of scale and location ($y = ax + b$, $a > 0$), and all assume that the outliers, if present, have means different from that of the other observations. Each of them treats a different generalization of the outlier problem. Murphy shows that the most powerful invariant procedure for selecting K observations as outliers is based on

$$[\alpha_n(x_{(n)}-\bar{x}) + \alpha_{n-1}(x_{(n-1)}-\bar{x}) + \cdots + \alpha_{n-k+1}(x_{(n-k+1)}-\bar{x})]/s \ .$$

His development assumes no external information as to σ^2, and that $\alpha_n > \alpha_{n-1} > \cdots > \alpha_{n-k+1}$ are known constants proportional to the amounts by which the means of the outliers differ from the mean of the non-outliers. Since the procedure depends on the α's, there is no procedure which is most powerful for all outlier means if $k > 1$. If the outliers are assumed to have the same distribution, however, the α's may be taken as equal, and an equivalent statistic is $(\bar{x}_1-\bar{x}_2)/s$, where $\bar{x}_1$ is the mean of the k highest observations and $\bar{x}_2$ the mean of the remaining observations. If the test is for only one outlier, the statistic is equivalent to $(x_{(n)}-\bar{x})/s$.

Paulson treats the case of the one-way classification with equal numbers in the cells. He develops a procedure for deciding that all cells have the same mean or that one cell has a mean higher than the others, the other means being equal. Paulson's procedure is based on $(x_{(n)}-\bar{x})/s$ when there is but one observation per cell, and the problem reduces to that of at most one outlier in a normal sample.

Kudō considers three groups of independently distributed observations. The first group is distributed $N(\mu,\sigma^2)$ except possibly for one outlier, which has a mean exceeding μ. The second group is $N(\mu,\sigma^2)$ and free of outliers. The third group is $N(\mu_1,\sigma^2)$ with $\mu_1 \neq \mu$. When there are no observations in the second group, the test statistic which Kudō derives reduces to that for which Quesenberry and David determine rejection constants. When there are observations only in the first group it reduces to $(x_{(n)}-\bar{x})/s$.

Dixon (1950) investigates the performance of outlier tests using random sampling. The tests studied include Grubbs' test and various tests based on ratios of differences of order statistics, such as $(x_{(n)}-x_{(n-1)})/(x_{(n)}-x_{(1)})$. Both scalar and location contamination are considered, with one or two outliers, or with some percent of the population sampled being contaminated. For one or two outliers, performance is the percentage of samples in which the outliers are extreme values and are detected as outliers. For contamination of a percentage of the population samples, performance is the percent of total contamination discovered.

David and Paulson (1965) mention several possible measures of the performance of outlier tests for the case of at most one outlier. In these measures the emphasis is on detection of outliers as an end in itself rather than a means of improving estimates of parameters. The different measures result from such considerations as whether the outlier is also the extreme value and whether the outlier is the only significant result in the sample. Some of these measures would be extremely difficult to evaluate numerically. For a variety of cases David and Paulson determine

$$\Pr \left\{ \frac{x_1 - \bar{x}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 + \nu s_\nu^2}} > V_\alpha^{(n,\nu)} \right\}$$

as a function of the shift in mean of x_1, which is assumed to be the only outlier, with $V_\alpha^{(n,\nu)}$ such that, with no outliers,

$$\Pr \left\{ \frac{x_{(n)} - \bar{x}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 + \nu s_\nu^2}} > V_\alpha^{(n,\nu)} \right\} = \alpha .$$

Thus they determine the probability that the outlier is significant, without regard to whether other values are also significant.

1.3 Problems Considered in This Study

In the present study problems of more than one outlier are considered, with emphasis on the case of two outliers. Usually we assume that a sample of independent observations $x_1, x_2, \ldots, x_n$ is such that

$$x_i \sim N(\mu, \sigma^2) \qquad i = 1,2,\ldots,j-1,j+1,\ldots,k-1,k+1,\ldots,n$$

$$x_j, x_k \sim N(\mu+\lambda, \sigma^2) \qquad \lambda > 0$$

$$j,k \text{ unknown}$$

i.e., x_j, x_k are outliers differing from the other observations in mean.

Some consideration is given to cases in which the two outliers do not have the same mean. In these cases it is assumed either that x_j and x_k have expectations $\mu+\lambda_1$, $\mu+\lambda_2$, with $\lambda_1, \lambda_2 > 0$, or $\mu+\lambda$, $\mu-\lambda$.

In Chapters 2, 3, and 4, respectively, the performance of the following three procedures for detecting outliers in this situation is evaluated:

1. Sequential test of maximum residual. In this procedure, $x_{(n)}$ is considered an outlier if

$$(x_{(n)}-\bar{x})/\sigma > c$$

when the variance is known, and if

$$(x_{(n)}-\bar{x})/s > c$$

when the variance is unknown. If $x_{(n)}$ is detected as an outlier, the test is repeated on the remaining n-1 observations. For two-sided alternatives $x_{(n)}-\bar{x}$ is replaced with the largest residual in absolute value.

2. Murphy's test. The two largest observations are considered outliers if

$$\frac{x_{(n)}-\bar{x} + x_{(n-1)}-\bar{x}}{s} > c \ .$$

If the variance is known, the denominator is replaced with σ.

3. Grubbs' test. The two largest observations are considered outliers if

$$\frac{\sum_{i=1}^{n-2} (x_{(i)}-\bar{x}_{n,n-1})^2}{\sum_{i=1}^{n} (x_i-\bar{x})^2} < c \qquad \text{where} \qquad \bar{x}_{n,n-1} = \sum_{i=1}^{n-2} x_{(i)}/n-2 \ .$$

Here c is a generic constant which provides in each case probability $1-\alpha$ that no observations will be declared outliers if in fact there are none.

These procedures are improvements over Thompson's procedure if there are multiple outliers. The sequential procedure has a desirable feature when the number of potential outliers is unknown, since the first step has good properties when there is only one outlier. Grubbs' and Murphy's tests are designed to test for exactly two outliers.

In evaluating performance, x_1 and x_2 are assumed to be outliers. The probabilities of rejection constants being exceeded for x_1 and x_2 were calculated, without regard to whether they are extreme values or other considerations. This is a direct generalization of the approach of David and Paulson.

For the sequential procedure there are three rather natural measures of performance as follows:

1. Probability of detecting at least one outlier,
2. Probability of detecting both outliers in two steps,
3. Probability that both outliers are significant at the first step.

For Grubbs' and Murphy's tests the choice is between two outliers and no outliers. The only measure considered is the probability that both x_1 and x_2 are detected. The performance measures were evaluated numerically as a function of λ/σ in a variety of cases.

Chapter 3 contains a development of the distribution of $(x_{(n)} + x_{(n-1)} - 2\bar{x})/\sigma$, and upper 5% and 1% percentage points. Also a simplified derivation is presented for an optimum procedure for detecting a predetermined number of outliers. This derivation is sufficiently general to include the procedures of Murphy, Paulson, and Kudō as special cases or easy extensions, but is simplified considerably by the use of sufficient statistics.

In Chapter 5 performances of the three procedures are compared, summarizing results of the preceding chapters. Finally, some conclusions as to the use of outlier tests are presented.

2. SEQUENTIAL TEST OF MAXIMUM RESIDUAL

2.1 Known Variance

2.1.1 The Test Procedure

Let $x_1, x_2, \ldots, x_n$ be a sample of independent observations, each distributed $N(\mu, \sigma^2)$ if there are no outliers. Outliers, if any, are distributed $N(\mu+\lambda, \sigma^2)$, $\lambda > 0$. If σ^2 is known we can assume for convenience that $\sigma^2 = 1$. We denote the ordered observations $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$. A suitable procedure for detection of a single outlier is to consider $x_{(n)}$ an outlier if

$$(2.1) \quad x_{(n)} - \bar{x} > V_\alpha^{(n)}$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ and $V_\alpha^{(n)}$ is the rejection constant tabulated by Grubbs (1950) such that

$$\Pr \{ x_{(n)} - \bar{x} > V_\alpha^{(n)} \} = \alpha$$

if actually there are no outliers. For at most one outlier, this procedure is optimum in a sense discussed in Chapter 3. In practice, however, the number of potential outliers is likely to be unknown. It seems reasonable to consider the sequential procedure of testing for a single outlier and, if an outlier is detected, rejecting it and repeating the test on the remaining data. Thus if (2.1) holds, we reject $x_{(n)}$ in the first stage. If also

$$x_{(n-1)} - \bar{x}_{(n)} > V_\alpha^{(n-1)}$$

where $\bar{x}_{(n)}$ is the mean of observations excluding $x_{(n)}$, we reject $x_{(n-1)}$, etc. until the test fails to reject an observation. In this way the probability of rejecting no observations if there are no outliers remains $1-\alpha$, while in some samples more than one observation is rejected.

2.1.2 Performance Criteria

To evaluate the performance of this procedure with two outliers we assume that the outliers are x_1 and x_2; i.e.

$$(2.2) \qquad x_i \sim N(\mu+\lambda,1) \qquad i = 1,2 \qquad \lambda > 0$$

$$x_i \sim N(\mu,1) \qquad i = 3,4,\ldots,n$$

and consider the following measures of performance:

P_a = Probability that at least one outlier is detected

P_b = Probability that both outliers are detected

P_c = Probability that both outliers are significant at the first stage .

Thus we consider only whether the test procedure yields significant results with the true outliers, without regard to whether other observations also are significant.

2.1.3 Calculation of Performance

The following easily verified results are helpful in numerical evaluation of the performance measures.

$$(2.3) \qquad V_\alpha^{(n-1)} < V_\alpha^{(n)}$$

$$(2.4) \qquad \text{If } x_1 > x_2, \text{ then } x_2-\bar{x}_1 > \frac{n}{n-1}(x_2-\bar{x}) .$$

$$(2.5) \qquad \text{If } x_1 > x_2 \text{ and } x_2-\bar{x} > V_\alpha^{(n)}, \text{ then } x_2-\bar{x}_1 > V_\alpha^{(n-1)} .$$

(2.6) The joint distribution of $x_1-\bar{x}$ and $x_2-\bar{x}$ is bivariate normal with $\rho = -\frac{1}{n-1}$.

(2.7) $x_i-\bar{x}$ and $x_j-\bar{x}_i$ are independent if $i \neq j$.

Here, as elsewhere, $\bar{x}_i$ denotes the mean of observations excluding x_i. That (2.3) holds can be verified from the table of $V_\alpha^{(n)}$ and can also be shown analytically for α of the usual size. Then (2.4) and (2.5) are simple algebraic results. The importance of (2.5) is that simultaneous first stage significance of x_1 and x_2 implies sequential significance. consequently

(2.8) $P_a > P_b > P_c$.

Calculation of the probabilities P_a, P_b, and P_c is now straightforward and proceeds as follows:

$$P_c = \Pr\{ \min(x_1-\bar{x}, x_2-\bar{x}) > V_\alpha^{(n)} \}$$

$$= \int_{V_\alpha^{(n)}}^{\infty} \int_{V_\alpha^{(n)}}^{\infty} f(y_1,y_2)\, dy_1 dy_2$$

where $f(y_1,y_2)$ is the bivariate normal density with

$$E(y_1) = E(y_2) = E(x_1-\bar{x}) = E(x_2-\bar{x}) = \frac{n-2}{n}\lambda$$

$$\sigma_{y_1}^2 = \sigma_{y_2}^2 = V(x_1-\bar{x}) = \frac{n-1}{n}$$

$$\rho = -\frac{1}{n-1} .$$

Also

$$P_b = Pr\{\, x_1 - \bar{x} > V_\alpha^{(n)},\; x_2 - \bar{x}_1 > V_\alpha^{(n-1)} \,\}$$

$$+ Pr\{\, x_2 - \bar{x} > V_\alpha^{(n)},\; x_1 - \bar{x}_2 > V_\alpha^{(n-1)} \,\} - P_c$$

$$= 2\, Pr\{\, x_1 - \bar{x} > V_\alpha^{(n)} \,\}\, Pr\{\, x_2 - \bar{x}_1 > V_\alpha^{(n-1)} \,\} - P_c$$

$$= 2 \left[1 - \Phi\left(\frac{V_\alpha^{(n)} - \frac{n-2}{n}}{\sqrt{\frac{n-1}{n}}} \right) \right] \left[1 - \phi\left(\frac{V_\alpha^{(n-1)} - \frac{n-2}{n-1}}{\sqrt{\frac{n-2}{n-1}}} \right) \right] - P_c$$

where

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{2\pi}\, e^{-\frac{t^2}{2}}\, dt$$

Similarly,

$$P_a = Pr\{\, x_1 - \bar{x} > V_\alpha^{(n)} \,\} + Pr\{\, x_2 - \bar{x} > V_\alpha^{(n)} \,\} - P_c$$

$$= 2 \left[1 - \Phi\left(\frac{V_\alpha^{(n)} - \frac{n-2}{n}}{\sqrt{\frac{n-1}{n}}} \right) \right] - P_c \;.$$

The bivariate normal integrals can be evaluated easily using tables published by National Bureau of Standards (1959) provided $\frac{1}{n-1}$ is an integral multiple of 0.05. Results for $n = 6$, 11, and 21 and $\alpha = .05$ and .01 are plotted in Figures 2.1(a)-(f). For comparison, the probability of detecting a single outlier is also shown. It appears that in most cases P_a is slightly higher than the probability of detecting a single outlier.

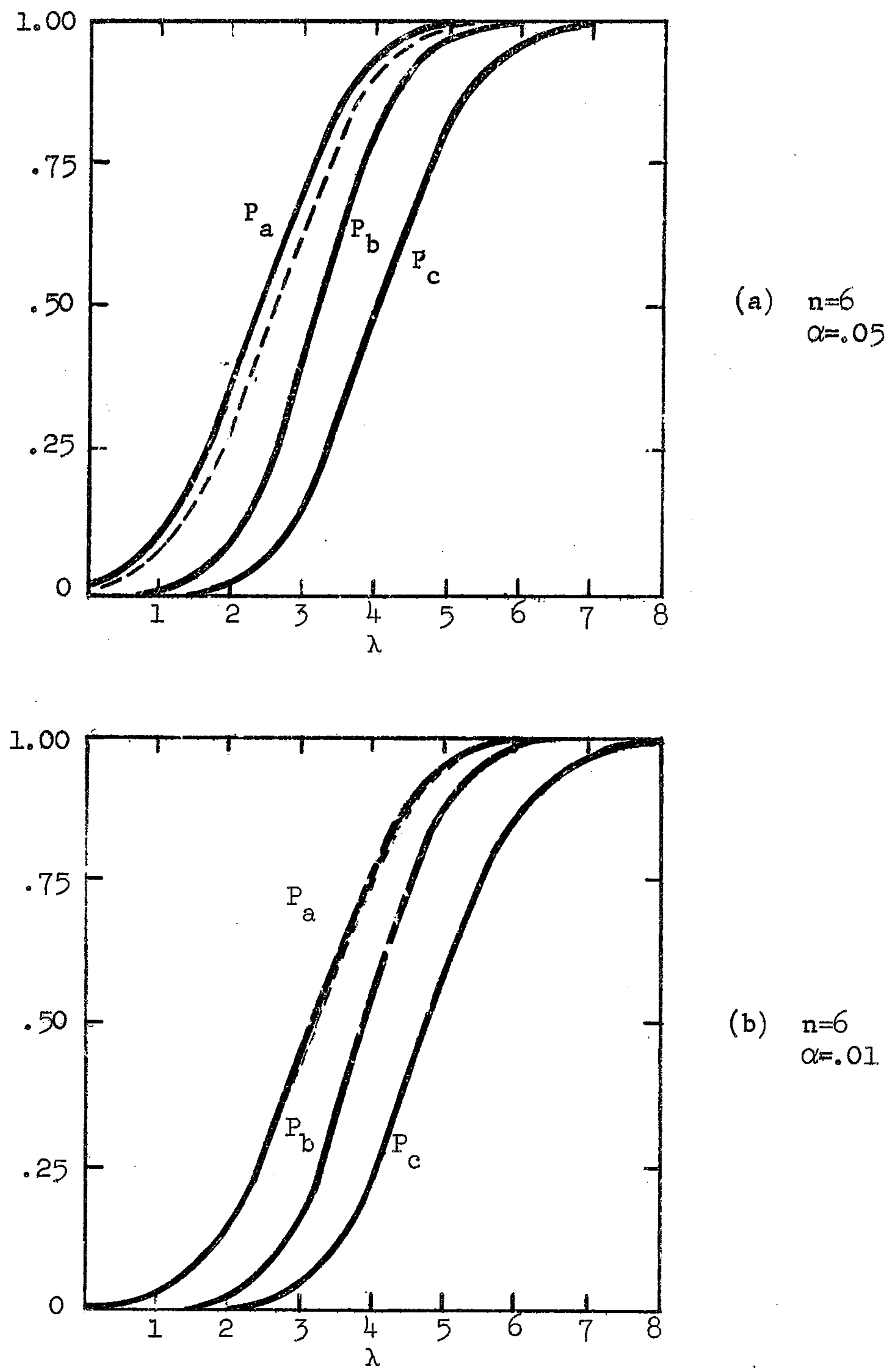
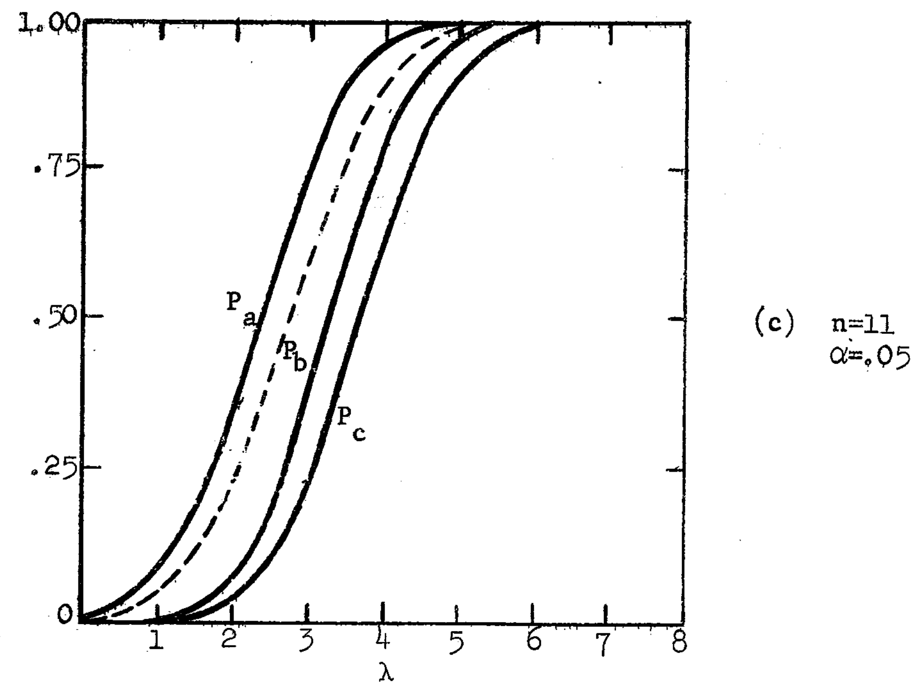


Figure 2.1 Performances P_a, P_b, P_c of sequential maximum residual test of section 2.1.1 at level α when $x_1, x_2 \sim N(\mu+\lambda, 1)$ and $x_3, \ldots, x_n \sim N(\mu, 1)$, with dashed lines showing the performance of the ordinary maximum residual test in the presence of one outlier



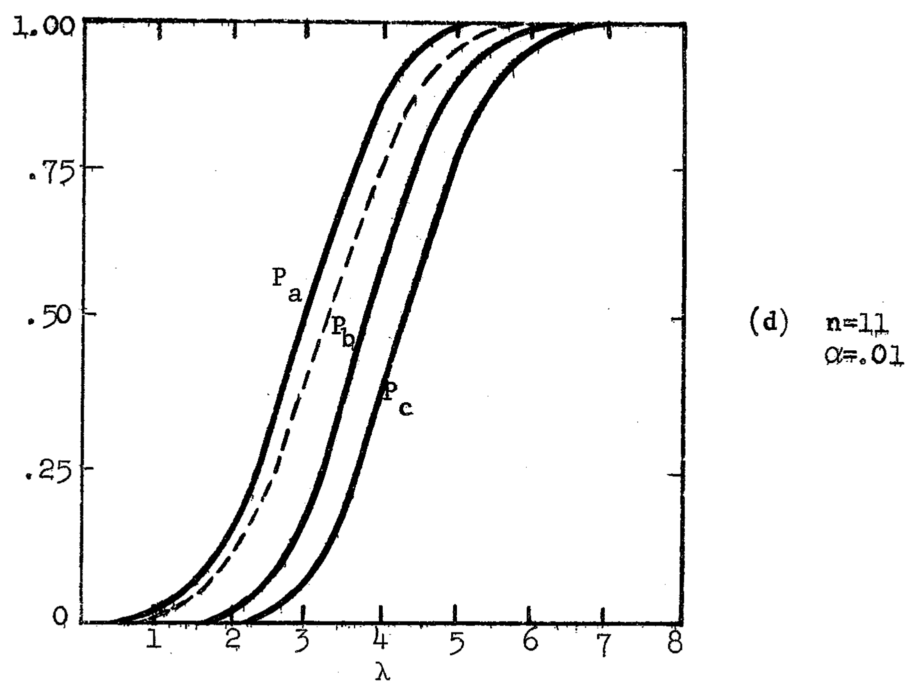


Figure 2.1 (continued)

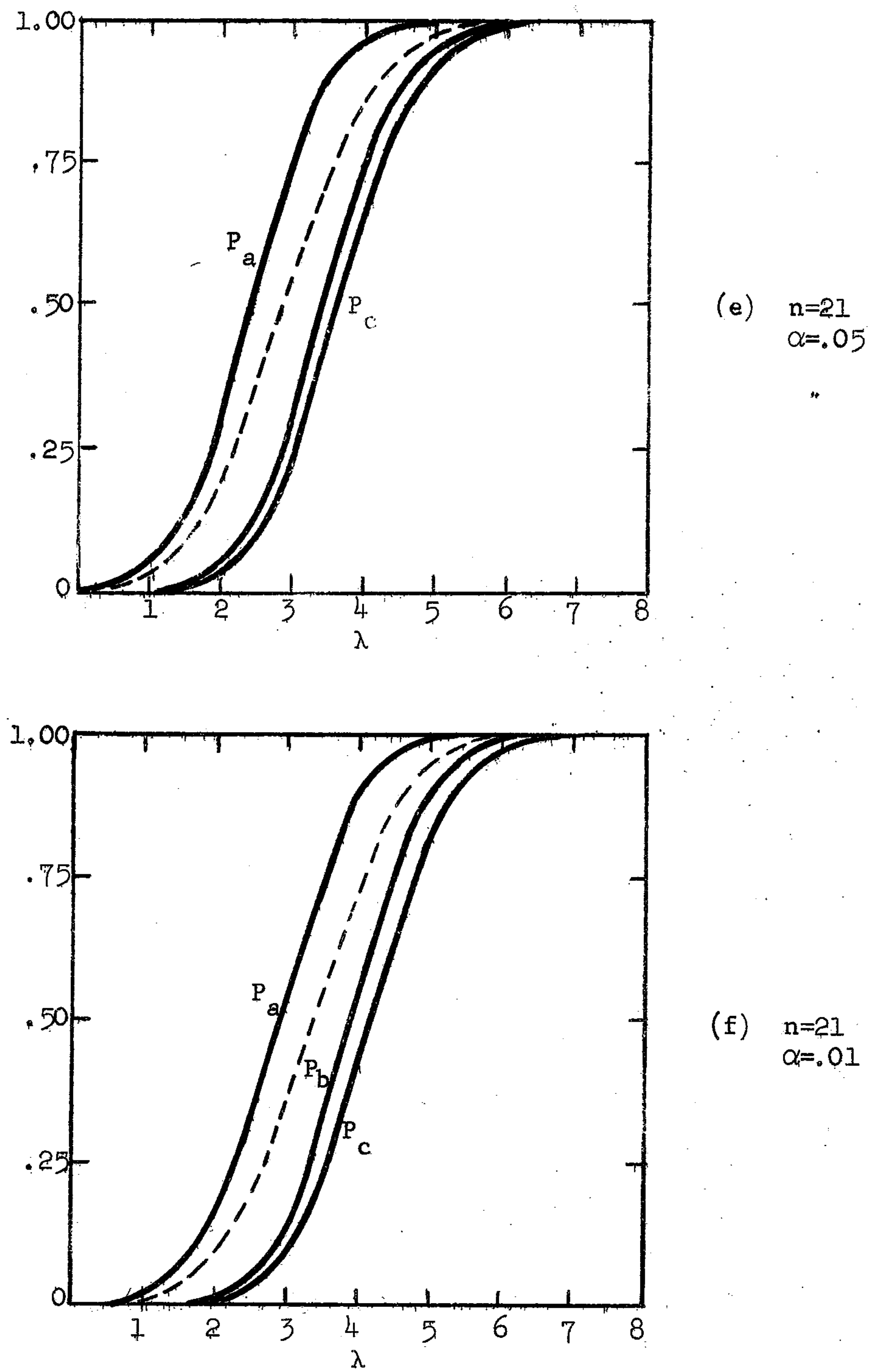


Figure 2.1 (continued)

2.1.4 Two-sided Alternatives

Up to now it has been assumed that the distribution of outliers is $N(\mu+\lambda,1)$, $\lambda > 0$. If we do not know the direction in which the distribution of the outliers has shifted, or if the two outlier distributions may have shifted in opposite directions, we consider the procedure of declaring outliers those observations x_i for which

$$(2.9) \qquad |x_i - \bar{x}| > W_\alpha^{(n)} .$$

Rejection constants, $W_\alpha^{(n)}$, for such tests were given by Halperin et al. (1955). As before, we consider the probabilities of significance of the true outliers. Now, however, we evaluate performance for the case

$$(2.10) \qquad \begin{aligned} x_1 &\sim N(\mu+\lambda,1) \qquad & \lambda > 0 \\ x_2 &\sim N(\mu-\lambda,1) & \\ x_i &\sim N(\mu,1) & i = 3,4,\ldots,n \end{aligned}$$

and take as performance measures

P_a = Probability of detecting at least one outlier

$$= \Pr \{ \max(|x_1-\bar{x}|, |x_2-\bar{x}|) > W_\alpha^{(n)} \}$$

P_c = Probability of detecting both outliers

$$= \Pr \{ \min(|x_1-\bar{x}|, |x_2-\bar{x}|) > W_\alpha^{(n)} \} .$$

In this case we do not assume the test to be repeated using observations not rejected using (2.9), since sequential rejection of two outliers in opposite directions implies first stage rejection of both. If

$$x_1 - \bar{x} > W_\alpha^{(n)}, \; x_2 - \bar{x}_1 < -W_\alpha^{(n-1)}$$

then

$$x_2 - \bar{x} < -W_\alpha^{(n-1)} + \bar{x}_1 - \bar{x} = -W_\alpha^{(n-1)} - \frac{x_1 - \bar{x}}{n-1} < -W_\alpha^{(n-1)} - \frac{W_\alpha^{(n)}}{n-1} .$$

This implies

$$x_2 - \bar{x} < -W_\alpha^{(n)}$$

if

$$-W_\alpha^{(n-1)} - \frac{W_\alpha^{(n)}}{n-1} < -W_\alpha^{(n)}$$

or

$$(2.11) \qquad W_\alpha^{(n-1)} > \frac{n-2}{n-1} W_\alpha^{(n)} .$$

That (2.11) holds can be verified numerically for all cases of interest. There is, of course, positive probability that an outlier is significant in the opposite direction from its true shift. If this happens, an additional outlier can be detected by repeating the test on the remaining observations. With the model (2.10), however, the probability that this occurs is negligible for other than small λ. We therefore disregard P_b, the probability of sequential detection of the two outliers.

Results for P_a and P_c using the model (2.9) are shown for some representative cases in Figures 2.2(a)-(f). The probabilities P_a and P_b for the one-sided model considered earlier are shown as dashed lines for comparison.

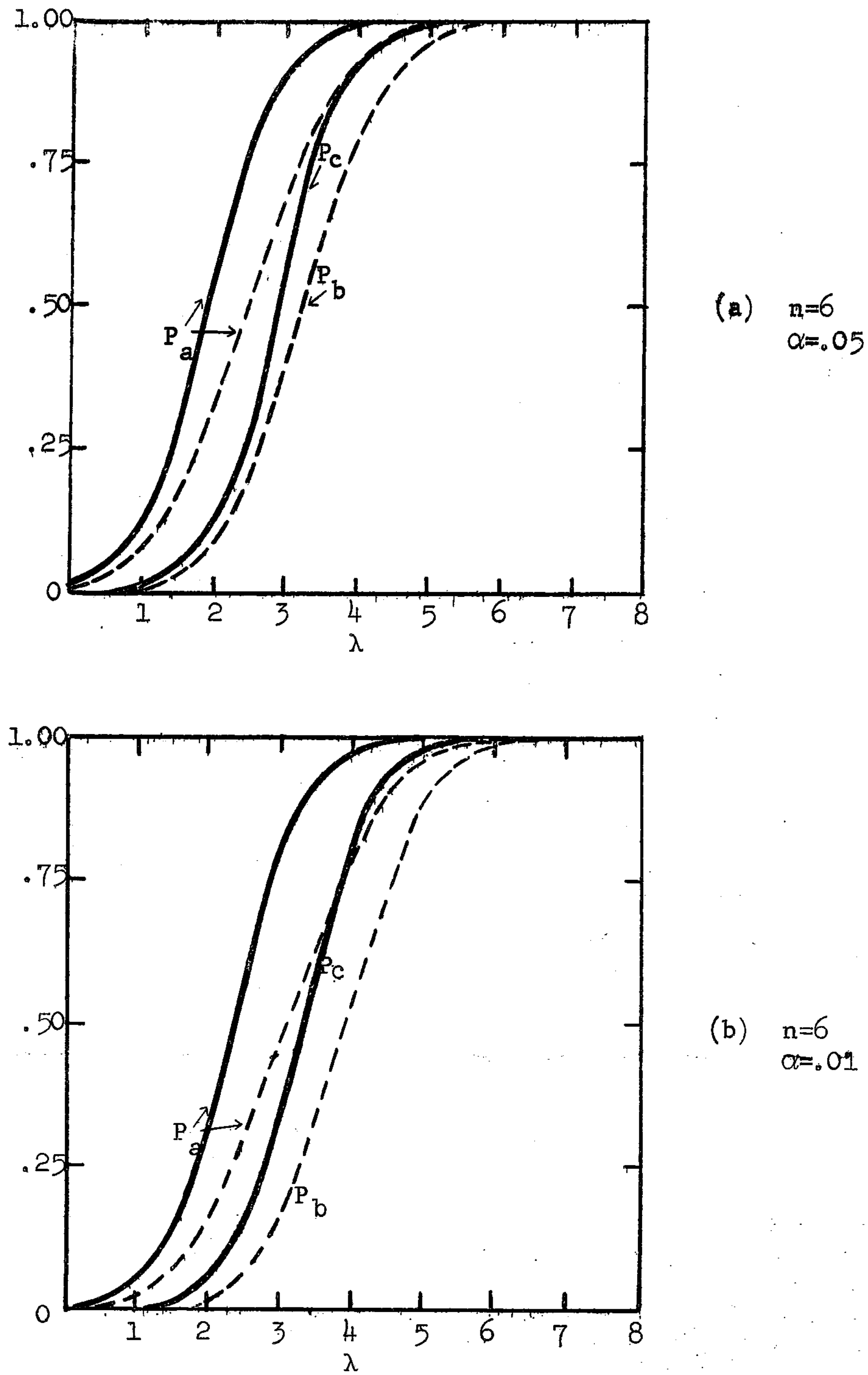


Figure 2.2 Performances P_a, P_c of sequential maximum residual test of section 2.1.4 when $x_1 \sim N(\mu+\lambda,1)$, $x_2 \sim N(\mu-\lambda,1)$ and $x_3,\ldots,x_n \sim N(\mu,1)$, with dashed lines showing P_a, P_b of the one-sided test of section 2.1.

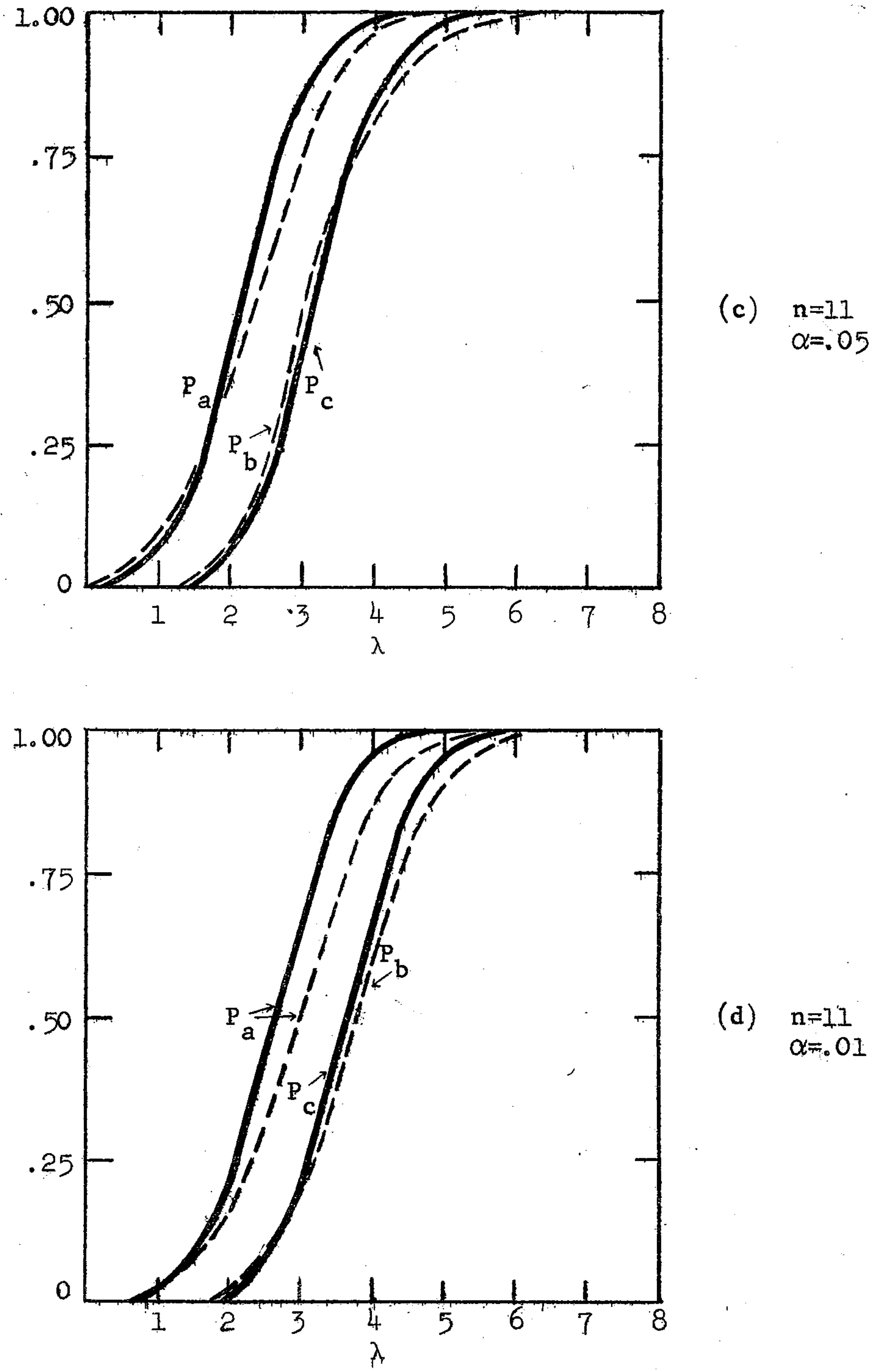
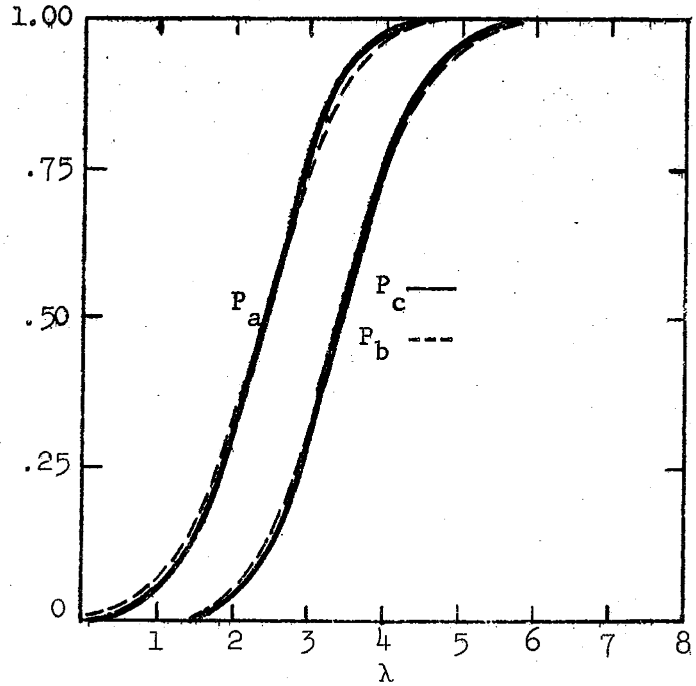
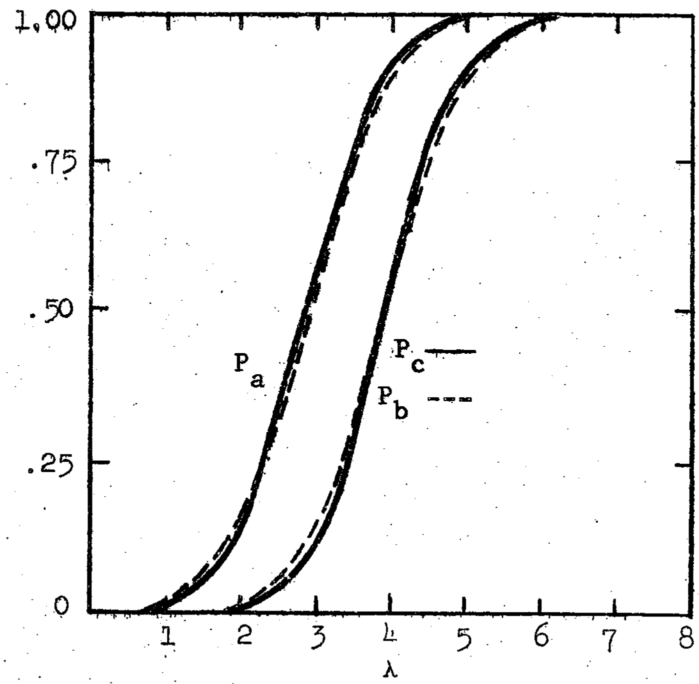


Figure 2.2 (continued)



(e) n=21
α=.05



(f) n=21
α=.01

Figure 2.2 (continued)

Performance for the model with both outliers on the same side with the direction of shift unknown was not evaluated. Except for small λ this situation is similar to the one-sided model (2.2) but using the larger rejection constants $W_\alpha^{(n)}$. The sequential procedure is again advantageous. The performance would be inferior to that of the one-sided case because of the larger rejection constants. On the other hand Figure 2.2 indicates that two outliers in opposite directions may be easier to detect than two on the same side even if the direction is known.

2.2 Unknown Variance

2.2.1 The Test Procedure

As with known variance, we consider a sample $x_1, x_2, \cdots, x_n$ of independent observations, each distributed $N(\mu, \sigma^2)$ if there are no outliers. Let

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + \nu s_\nu^2 ,$$

$$S_K^2 = \sum_{i \neq K} (x_i - \bar{x}_K)^2 + \nu s_\nu^2$$

where s_ν^2 is a mean square estimate of variance based on ν d.f., independent of $x_1, x_2, \cdots, x_n$. We consider a sequential procedure in which we reject the largest observation in the first stage if

$$\frac{x_{(n)} - \bar{x}}{S} > V_\alpha^{(n,\nu)}$$

where $V_\alpha^{(n,\nu)}$ is the rejection constant tabulated by Quesenberry and

David (1961). Thus the first stage is based on the statistic proposed by Kudō (1956), which is optimum in a particular sense if there is at most one outlier. If $x_{(n)}$ is declared an outlier, the test is repeated with the remaining observations, i.e. if

$$\frac{x_{(n-1)} - \bar{x}_{(n)}}{s_{(n)}} > V_{\alpha}^{(n-1,\nu)}$$

then $x_{(n-1)}$ is also declared an outlier.

2.2.2 Performance Criteria

We now assume that the only outliers in the sample are x_1 and x_2 and consider probabilities that these observations yield significance in the relations above, but without regard to order relationships. We use the performance measures defined in section 2.1.2.

2.2.3 Calculation of Performance

The results (2.3) - (2.5) still hold. In addition we need

(2.12) $$V_{\alpha}^{(n-1,\nu)} < \frac{n}{n-1} V_{\alpha}^{(n,\nu)}$$

(2.13) $$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{n}{n-1} (x_K - \bar{x})^2 + \sum_{i \neq K} (x_i - \bar{x}_K)^2$$

(2.14) If $x_1 - \bar{x} > x_2 - \bar{x} > V_{\alpha}^{(n,\nu)} S$, then $x_2 - \bar{x}_1 > V_{\alpha}^{(n-1,\nu)} S_1$.

(2.15) $\frac{x_1 - \bar{x}}{S}$ and $\frac{x_2 - \bar{x}_1}{S_1}$ are independent.

The inequality (2.12) is verified numerically from the table of rejection constants. Relation (2.13) is an algebraic identity; it is frequently useful that the two terms on the right are independent. To prove (2.14), we have from (2.4) and (2.13)

$$x_2-\bar{x}_1 > \frac{n}{n-1}(x_2-\bar{x}) > \frac{n}{n-1} v_\alpha^{(n,\nu)} s$$

$$= \frac{n}{n-1} v_\alpha^{(n,\nu)} \sqrt{\frac{n}{n-1}(x_1-\bar{x})^2 + \sum_{i=2}^{n}(x_i-\bar{x}_1)^2 + \nu s_\nu^2}$$

$$> \frac{n}{n-1} v_\alpha^{(n,\nu)} s_1 > v_\alpha^{(n-1,\nu)} s_1$$

the last step following from (2.12). This means that, as with known variance, simultaneous first stage significance of two observations implies two-stage significance. Hence relation (2.8) also holds for unknown variance. Conclusion (2.15) was proved by Quesenberry and David using a theorem of Basu. They needed the conclusion only in the case of $\lambda = 0$, but it follows immediately for $\lambda \neq 0$.

Evaluation of the performance measures is more difficult with unknown variance, but can be accomplished by numerical integration. Before proceeding with this, it is helpful to recall that under some conditions $P_c \equiv 0$. Pearson and Chandra Sekar (1936) showed that

$$(2.16) \qquad \frac{x_{(n-1)}-\bar{x}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \leq \sqrt{\frac{n-2}{2n}}$$

and

$$(2.17) \qquad \frac{|x-\bar{x}|_{(n-1)}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \leq \sqrt{\frac{1}{2}}$$

where $|x-\bar{x}|_{(n-1)}$ denotes the second largest of the residuals in absolute value. Thus we also have

$$(2.18) \quad \frac{x_{(n-1)}-\bar{x}}{S} \leq \sqrt{\frac{n-2}{2n}}$$

which is to say that for any n and ν, one can choose a rejection constant so large that the simultaneous rejections of two outliers is impossible. (The condition for $P_c \equiv 0$ is implicit in expression 3.3 of Quesenberry and David.) For $\alpha = 0.05$ and 0.01, the n and ν for which $P_c \equiv 0$ are as follows:

n	ν	
	α=.05	α=.01
3	$\leq$ 15	$\leq$ 24
4	$\leq$ 10	$\leq$ 18
5	$\leq$ 8	$\leq$ 15
6	$\leq$ 7	$\leq$ 12
7	$\leq$ 6	$\leq$ 11
8	$\leq$ 5	$\leq$ 10
9	$\leq$ 4	$\leq$ 9
10	$\leq$ 3	$\leq$ 8
12	$\leq$ 1	$\leq$ 6
14	0	$\leq$ 4
15		$\leq$ 3
21		0

For $n > 14$ with $\alpha = .05$ and $n > 21$ with $\alpha = .01$, P_c is positive even if $\nu = 0$.

To calculate the performance measures, we consider first the probability that x_1 is significant at the first stage, i.e.:

$$\Pr\{\ x_1-\bar{x} > V_{\alpha}^{(n,\nu)}S\ \} = \Pr\{\ (x_1-\bar{x})^2 > V^2S^2,\ x_1-\bar{x} > 0\ \}$$

$$= \Pr\{\ (x_1-\bar{x})^2 > V^2\ [\tfrac{n}{n-1}\ (x_1-\bar{x})^2 + \sum_{i=2}^{n}(x_i-\bar{x}_1)^2 + \nu s_\nu^2],\ x_1-\bar{x} > 0\ \}$$

$$= \Pr\{\ (x_1-\bar{x})^2 > V^2\ [\tfrac{n}{n-1}\ (x_1-\bar{x})^2 + \tfrac{n-1}{n-2}\ (x_2-\bar{x}_1)^2 + \sum_{i=3}^{n}(x_i-\bar{x}_{12})^2$$

$$+ \nu s_\nu^2\ ],\ x_1-\bar{x} > 0\ \}$$

where $\bar{x}_{12}$ is the mean of observations excluding x_1 and x_2. When there is no danger of confusion, we write for $V_{\alpha}^{(n,\nu)}$ simply V. In the last expression the four terms in square brackets are mutually independent. The sum of the last two of these terms is distributed as χ^2 with $n-3+\nu$ d.f. Letting

$$x_1-\bar{x} = t_1, \qquad x_2-\bar{x} = t_2\ ,$$

we have

$$\Pr\{\ x_1-\bar{x} > VS\ \} = \Pr\{\ t_1^2 > V^2\ [\tfrac{n}{n-1}\ t_1^2 + \tfrac{n-1}{n-2}\ (t_2 + \tfrac{t_1}{n-1})^2$$

$$+ \chi^2_{n-3+\nu}],\quad t_1 > 0\ \}$$

which, after some rearrangement, becomes

$$\text{(2.19)}\qquad \Pr\{\ x_1-\bar{x} > VS\ \} = \Pr\{\ t_1^2\ [\tfrac{n-2}{V^2} - (n-1)] - 2t_1t_2 - (n-1)t_2^2$$

$$> (n-2)\ \chi^2_{n-3+\nu},\quad t_1 > 0\ \}\ .$$

The discriminant of the quadratic function is

$$4 + 4 \left[\frac{n-2}{V^2} - (n-1)\right](n-1)$$

which is positive, since $V^2 < \frac{n-1}{n}$. Consequently, for fixed χ^2 the region in the t_1, t_2 plane where x_1 is significant is bounded by one branch of a hyperbola. By analogy,

$$(2.20) \quad \Pr\{ x_2 - \bar{x} > VS \} = \Pr\{ t_2^2 \left[\frac{n-2}{V^2} - (n-1)\right] - 2t_1 t_2 - (n-1) t_1^2 > (n-2)\, \chi^2_{n-3+\nu},\ \ t_2 > 0 \}.$$

In Figure 2.3 the situation for $n = 11$, $\nu = 0$, $\alpha = .05$ is represented. The quadratic expressions in (2.19) and (2.20) specify families of hyperbolas intersecting the t_1 and t_2 axes, respectively. Each family is indexed by $\chi^2_{n-3+\nu}$. The hyperbolas drawn are for $\chi^2_{.10}$, $\chi^2_{.50}$, and $\chi^2_{.90}$, where χ^2_p is the solution of $\Pr\{ \chi^2 < \chi^2_p \} = p$. Asymptotes, which correspond to $\chi^2 = 0$, are also shown.

For fixed χ^2 the probability that x_1 is significantly high is the integral of the joint density of t_1 and t_2 over the region bounded by the branch intersecting the positive t_1 axis of the hyperbola corresponding to the fixed χ^2. The probabilities that x_1 is significantly low and that x_2 is significantly high or low can be obtained in the same way, still conditioned on χ^2. The unconditional probabilities can be obtained by integrating numerically the conditional probabilities with respect to the pdf of χ^2.

In this example the probability of simultaneous rejection of x_1 and x_2 is zero, which is seen from the non-overlapping of the hyperbolas for

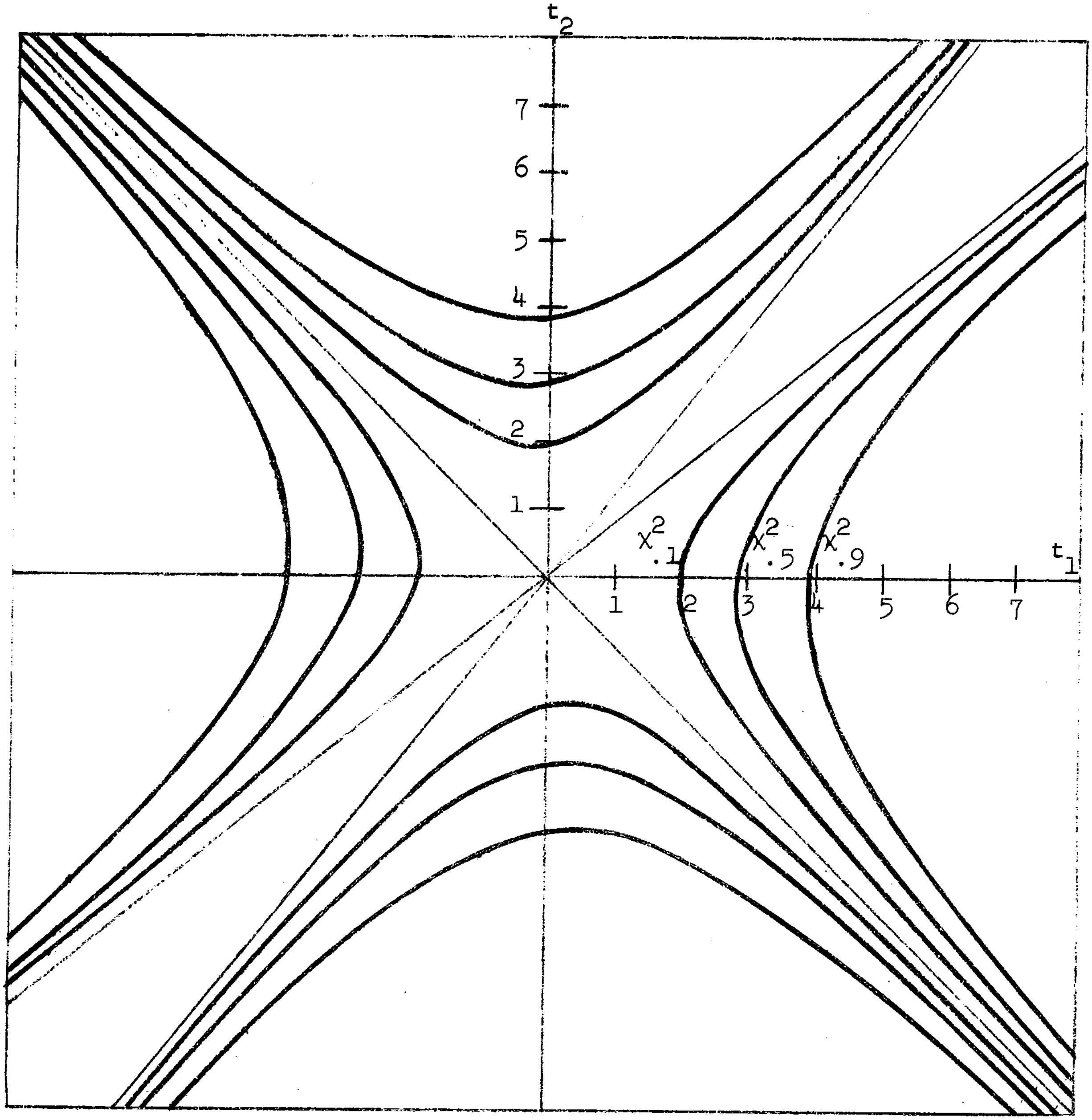


Figure 2.3 Rejection regions for first stage of sequential maximum residual test of section 2.2.1 for $n = 11$, $\nu = 0$, $\alpha = .05$, and χ^2 fixed, showing corresponding regions for rejection of low outliers

all χ^2. Otherwise the probability of simultaneous rejection, P_c, can be computed by integrating over the overlapping regions. Also, in this example, the two asymptotes in the second and fourth quadrants virtually coincide. This is not generally true. In fact, the probability of simultaneous rejection of one high and one low observation can be obtained by integrating over the appropriate overlapping regions in those quadrants. Other examples, in which simultaneous rejection has non-zero probability, are shown in Figure 2.4. In these figures only the positive branches of the hyperbolas for $\chi^2_{.50}$ are drawn.

Once the probabilities of first stage rejection of one or both of x_1 and x_2 are established, calculation of P_a and P_b is relatively easy. We have

$$P_a = \Pr\{ x_1-\bar{x} > V_\alpha^{(n,\nu)}S \} + \Pr\{ x_2-\bar{x} > V_\alpha^{(n,\nu)}S \}$$

$$- \Pr\{ \min(x_1-\bar{x},x_2-\bar{x}) > V_\alpha^{(n,\nu)}S \}$$

$$= 2 \Pr\{ x_1-\bar{x} > V_\alpha^{(n,\nu)}S \} - P_c$$

$$P_b = \Pr\{ x_1-\bar{x} > V_\alpha^{(n,\nu)}S \} \Pr\{ x_2-\bar{x}_1 > V_\alpha^{(n-1,\nu)}S_1 \}$$

$$+ \Pr\{ x_2-\bar{x} > V_\alpha^{(n,\nu)}S \} \Pr\{ x_1-\bar{x}_2 > V_\alpha^{(n-1,\nu)}S_2 \} - P_c$$

$$= 2 \Pr\{ x_1-\bar{x} > V_\alpha^{(n,\nu)}S \} \Pr\{ x_2-\bar{x}_1 > V_\alpha^{(n-1,\nu)} \} - P_c .$$

Probabilities of rejecting one outlier after the other has been rejected, such as $\Pr\{ x_2-\bar{x}_1 > V_\alpha^{(n-1,\nu)}S_1 \}$, are the only problems remaining. These probabilities can be expressed as noncentral t integrals and

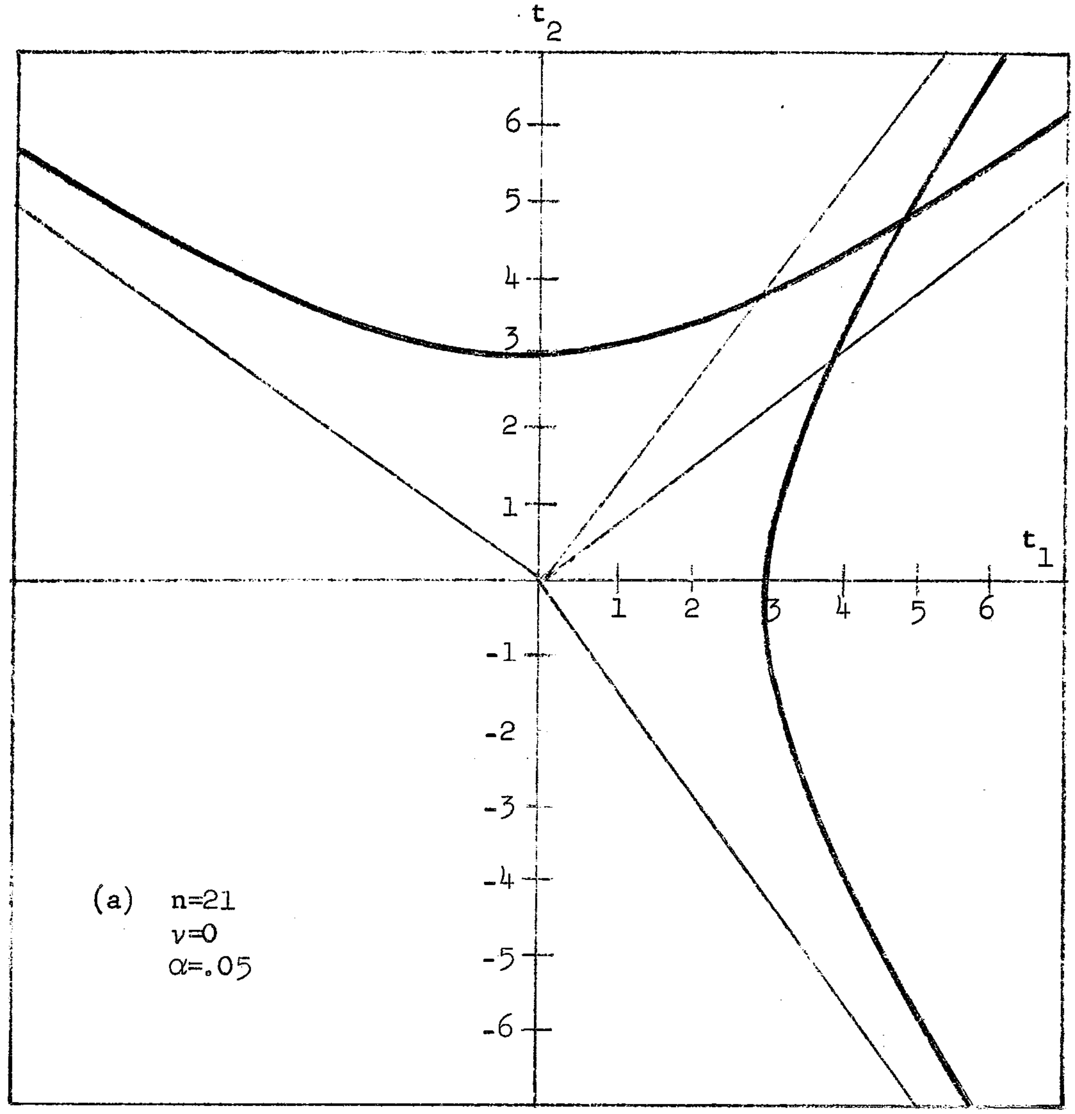


Figure 2.4 Rejection regions for first stage of sequential maximum residual test of section 2.2.1 for $\chi^2 = \chi^2_{.5}$

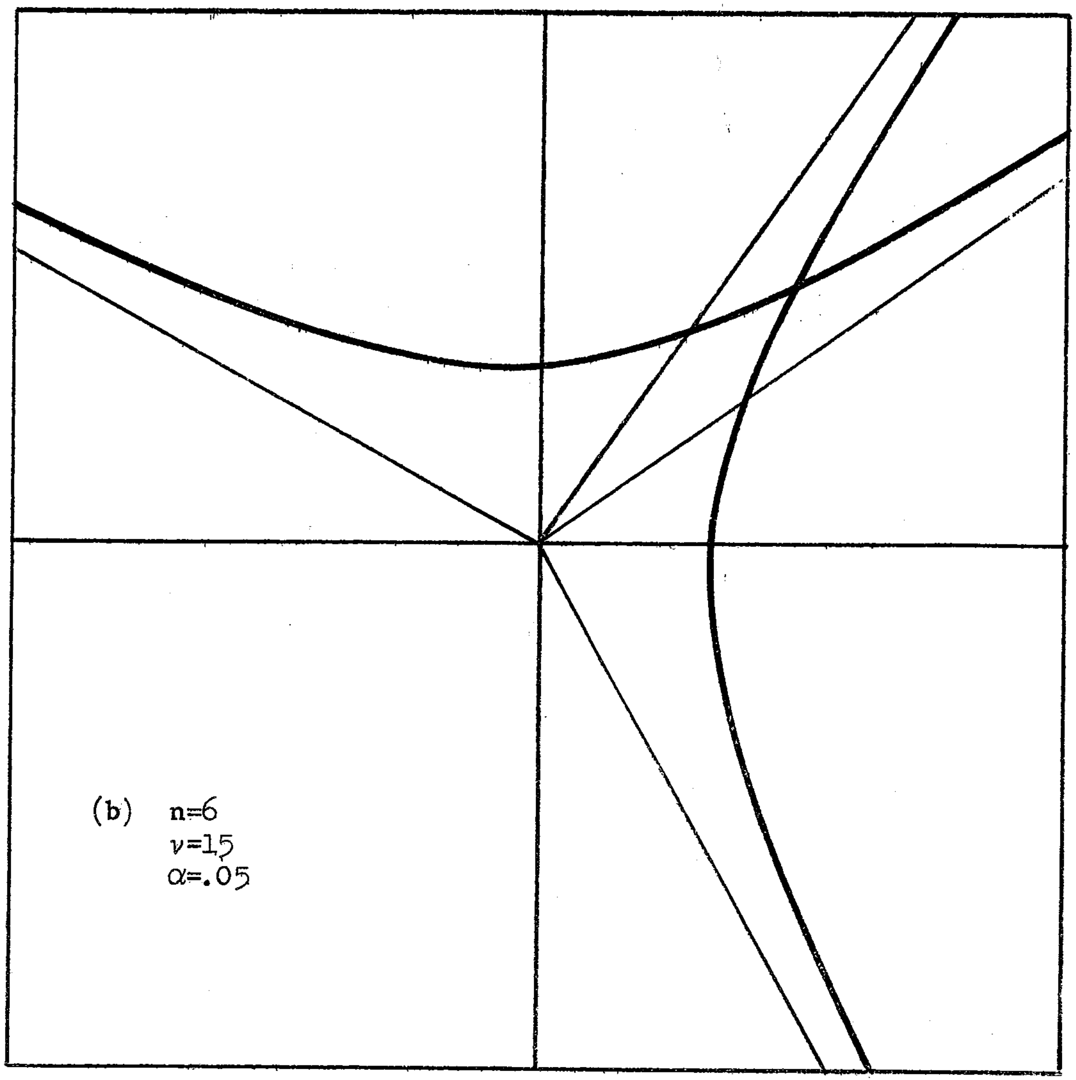


Figure 2.4 (continued)

evaluated from tables as did David and Paulson (1965), or obtained by numerical integration. In fact they can be obtained with the numerical integration scheme just discussed for first stage rejection probabilities, if the means of the bivariate normal variables t_1 and t_2 are properly adjusted. With sample size n-1 and one outlier, the means of t_1 and t_2 can be taken as $\frac{n-2}{n-1}\lambda$ and 0. This is an unnecessarily complicated approach, but provides some useful checks on numerical results.

The computational procedure for first stage rejection probabilities is also applicable if there are two outliers with unequal means. In this case, of course, the last formula for P_b, above, no longer holds.

The geometric representation used here verifies the results of Pearson and Chandra Sekar. Because of symmetry, the condition that hyperbolas do not overlap in the first (or third) quadrant can be expressed by requiring the slope γ_o of the t_1 asymptote to be less than one, i.e.

$$\gamma_o = \frac{-1 + \sqrt{1 + (n-1)[\frac{n-2}{V^2} - (n-1)]}}{n-1} < 1$$

which implies

$$V^2 > \frac{n-2}{2n}$$

verifying (2.16). Also, requiring the slope, γ_1, of the asymptote in the fourth (or second) quadrant to exceed -1, i.e.

$$\gamma_1 = \frac{-1 - \sqrt{1 + (n-1)[\frac{n-2}{V^2} - (n-1)]}}{n-1} > -1$$

yields

$$v^2 > \frac{1}{2}$$

verifying (2.17).

2.2.4 Asymptotic Properties of Performance Criteria

If the two outliers are from the same distribution, the mean of the bivariate normal distribution of t_1 and t_2 is on the line $t_1 = t_2$, specifically at the point $(\frac{n-2}{n}\lambda, \frac{n-2}{n}\lambda)$. Consequently if $P_c > 0$, i.e. if the hyperbolas for first stage rejection overlap, then $P_c \longrightarrow 1$ as $\lambda \longrightarrow \infty$. But if $P_c = 0$, the probability of detecting even one outlier approaches 0 as $\lambda \longrightarrow \infty$, except in the case in which the asymptotes coincide, for which $P_a \longrightarrow 1$.

If the means of the two outliers are different, say λ_1 and λ_2, let

$$\eta = \lambda_2/\lambda_1 , \qquad \gamma = \frac{Et_2}{Et_1} = \frac{(n-1)\eta-1}{n-1-\eta} .$$

Then, as λ_1 and λ_2 increase, the mean of the distribution of t_1 and t_2 moves along the line $t_2 = \gamma t_1$. If this line does not intersect the hyperbolas of either family, the probability of rejecting either outlier approaches zero as $\lambda_1, \lambda_2 \longrightarrow \infty$. The portion of the λ_1, λ_2 plane for which detection of either outlier becomes increasingly difficult as λ_1, λ_2 increase can be expressed as the region

$$\gamma_o < \gamma < \frac{1}{\gamma_o}$$

provided $\gamma_o < 1$. If $\gamma_o > 1$, there is no such region. Some representative values of γ_o are as follows:

n	ν	$V_{.05}$	γ_o
6	0	.815	.474
	2	.732	.662
	4	.666	.826
	6	.614	.971
11	0	.706	.803
	5	.600	1.128
	10	.528	1.396
	15	.477	1.622

The value of γ_o depends only on n and V, not specifically on ν, although for given α, V depends on ν.

2.2.5 Numerical Results

The foregoing discussion provides some indication of the circumstances in which this sequential procedure might be effective in a sample containing two outliers. A rough indication of what is likely to happen at the first stage for any λ_1 and λ_2 can be seen from the families of hyperbolas. Numerical results for P_a, P_b, and P_c have been obtained for a few cases and are shown in the following figures.

Figure 2.5 shows performance characteristics for n = 11 and n = 21 for unknown variance with $\epsilon = 0$, $\alpha = .05$, $\lambda_1 = \lambda_2$. In the first case simultaneous rejection of x_1 and x_2 is impossible, and the performance characteristics approach zero as λ increases.

Figure 2.6 is also for n = 11, but here $\lambda_1 = 2\lambda_2$. Again simultaneous rejection of two outliers is impossible, but P_a and P_b approach 1 as λ increases. Performance for known variance in this case is also shown.

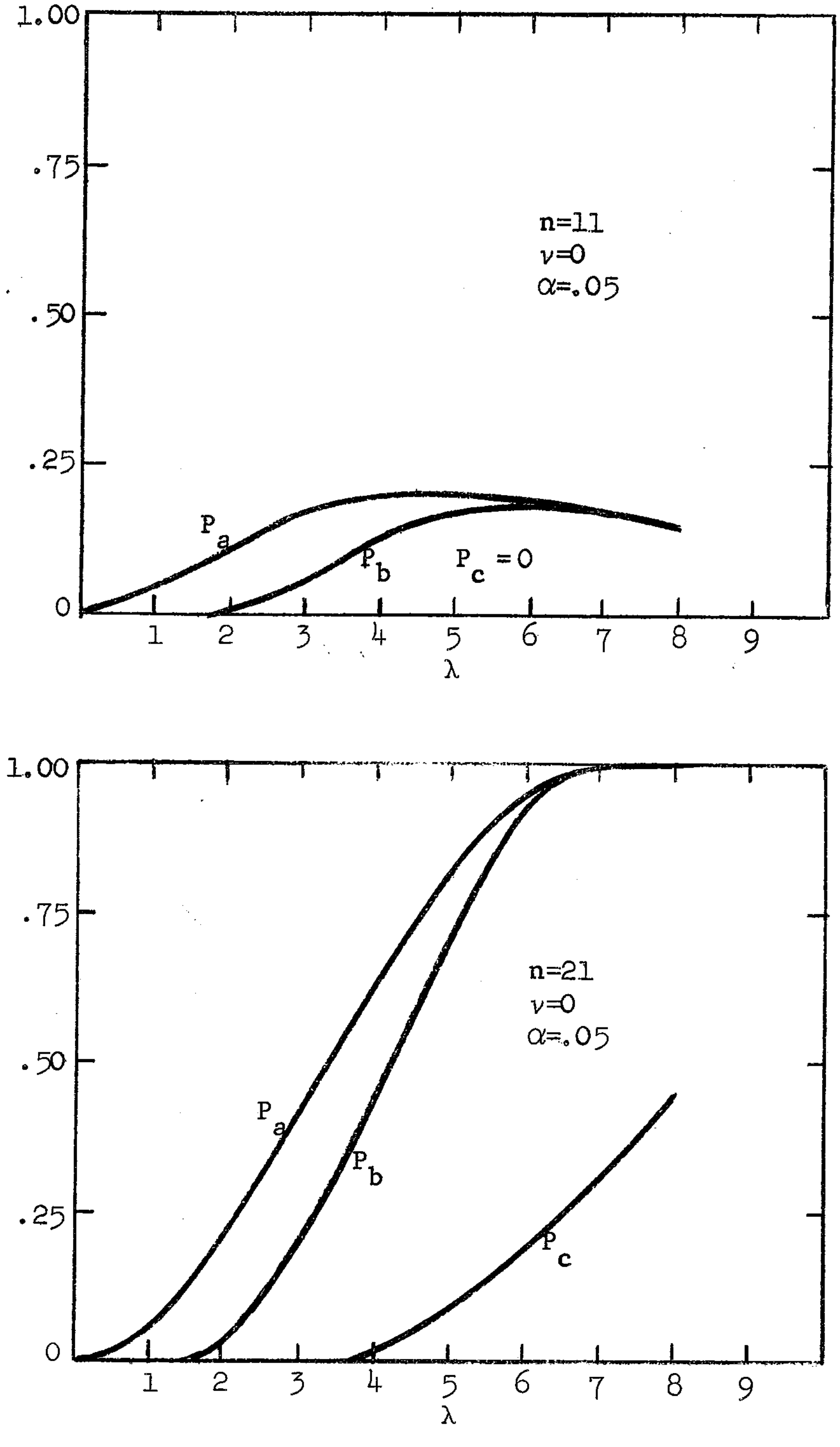


Figure 2.5 Performances P_a, P_b, P_c of sequential maximum residual test of section 2.2.1 when α = .05 and $x_1, x_2 \sim N(\mu+\lambda, \sigma^2)$, $x_3, \ldots, x_n \sim N(\mu, \sigma^2)$

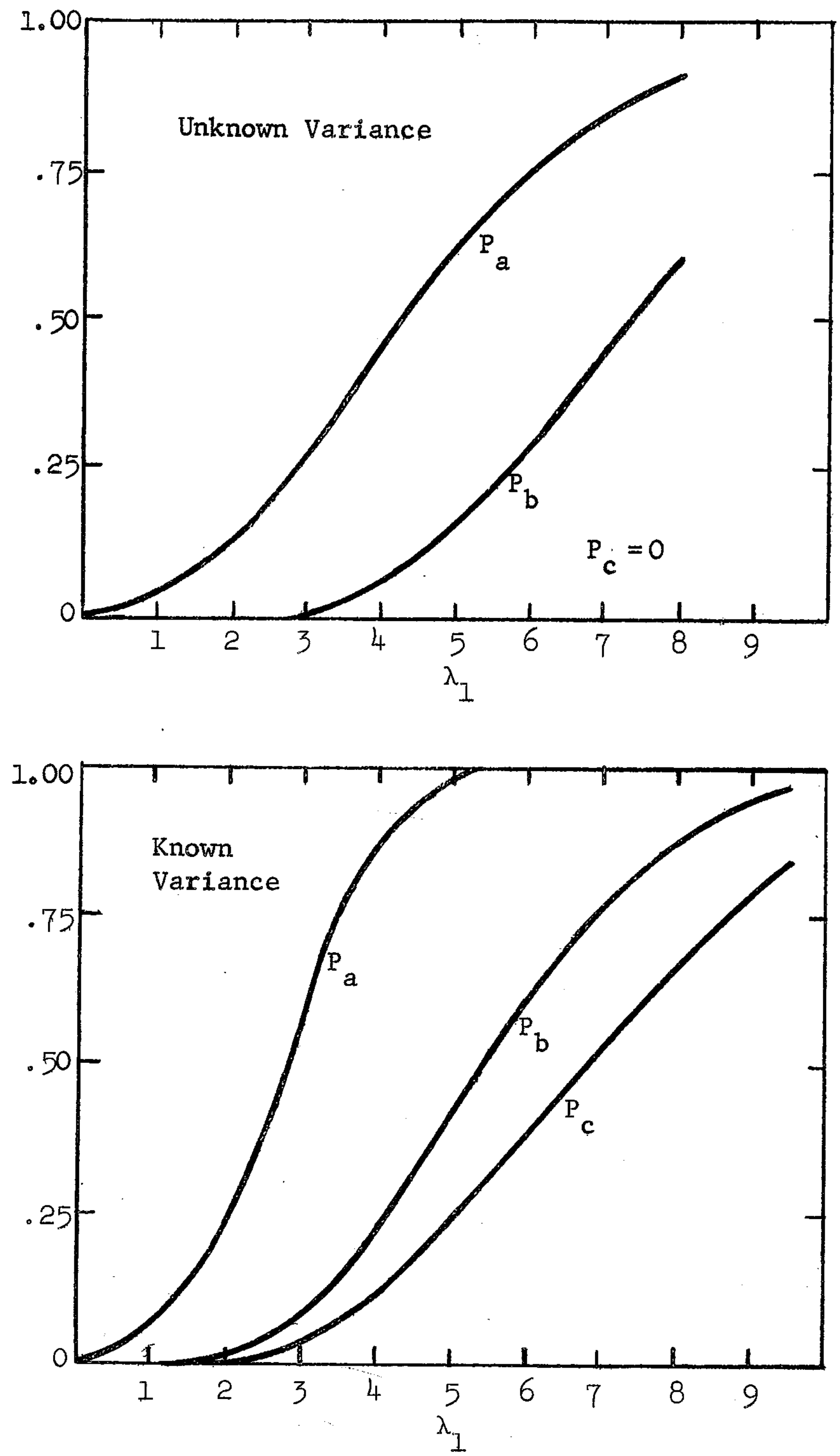


Figure 2.6 Performances P_a, P_b, P_c of sequential maximum residual test of sections 2.2.1 and 2.1.1 when $\alpha = .05$ and $x_1 \sim N(\mu+\lambda_1, \sigma^2)$, $x_2 \sim N(\mu + \frac{1}{2}\lambda_1, \sigma^2)$, $x_3, \ldots, x_n \sim N(\mu, \sigma^2)$

Figure 2.7 shows a comparison of the value of internal and external d.f. with respect to the performance characteristics, with $\lambda_1 = \lambda_2$. It appears that external d.f. are not necessarily more valuable in detecting outliers.

2.2.6 Two-sided Alternatives

We consider here the problem with unknown variance of testing for outliers which may have shifted by different amounts, or even in different directions. The first stage of the procedure is to class as an outlier the observation with the largest absolute residual if

$$\frac{|x-\bar{x}|_n}{S} > W_\alpha^{(n,\nu)}$$

where $W_\alpha^{(n,\nu)}$ is the two-sided rejection constant tabulated by Quesenberry and David. If an observation is classed as an outlier at this stage, the test is repeated on the remaining observations, etc. Unlike the corresponding test with known variance, this procedure can reject an observation in the second stage in the opposite direction from one rejected in the first stage. In fact simultaneous detection at the first stage ordinarily implies sequential detection. An explanation of conditions for which this is true is given in Appendix 7.1.

Earlier it was pointed out that there are two families of hyperbolas which relate to the efficiency of the maximum studentized residual test with two outliers shifted to the right. This is also true with one high and one low outlier. For means shifted by the same amount, λ, in opposite directions, the probability that both outliers are detected at the first stage approaches one as $\lambda \longrightarrow \infty$ if the two families overlap

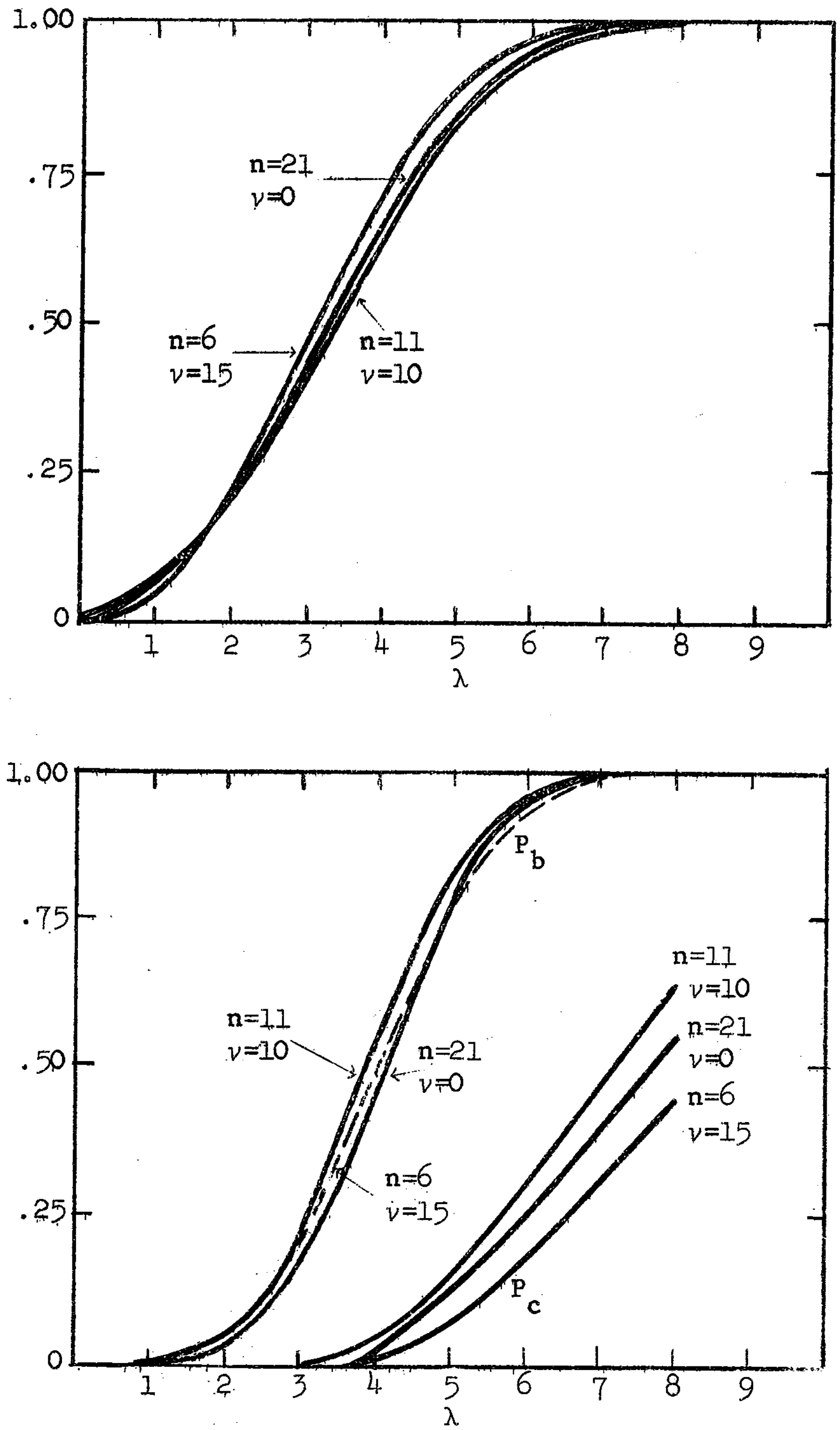


Figure 2.7 Performances P_a, P_b, P_c of sequential maximum residual test of section 2.2.1 when $n+\nu = 21$, $\alpha = .05$ and $x_1, x_2 \sim N(\mu+\lambda, \sigma^2)$, $x_3, \ldots, x_n \sim N(\mu, \sigma^2)$

in the second and fourth quadrants. The condition for the hyperbolas to overlap is $\gamma_1 < -1$, or $W^2 < \frac{1}{2}$. If $W^2 > \frac{1}{2}$, the probability of detecting even one outlier approaches 0 as $\lambda \to \infty$. The portion of the λ_1, λ_2 plane for which the first stage detection probabilities approach 0 for $\lambda_1, \lambda_2 \to \infty$, λ_2/λ_1 constant is obtained from

$$\frac{1}{\gamma_1} < \gamma < \gamma_1 .$$

Some representative values of γ_1 for the two-sided situation are as follows:

n	ν	$W_{.05}$	γ_1
6	0	.844	-.604
	2	.771	-.821
	4	.708	-.997
	6	.657	-1.145
11	0	.744	-.897
	5	.638	-1.205
	10	.566	-1.449
	15	.513	-1.659

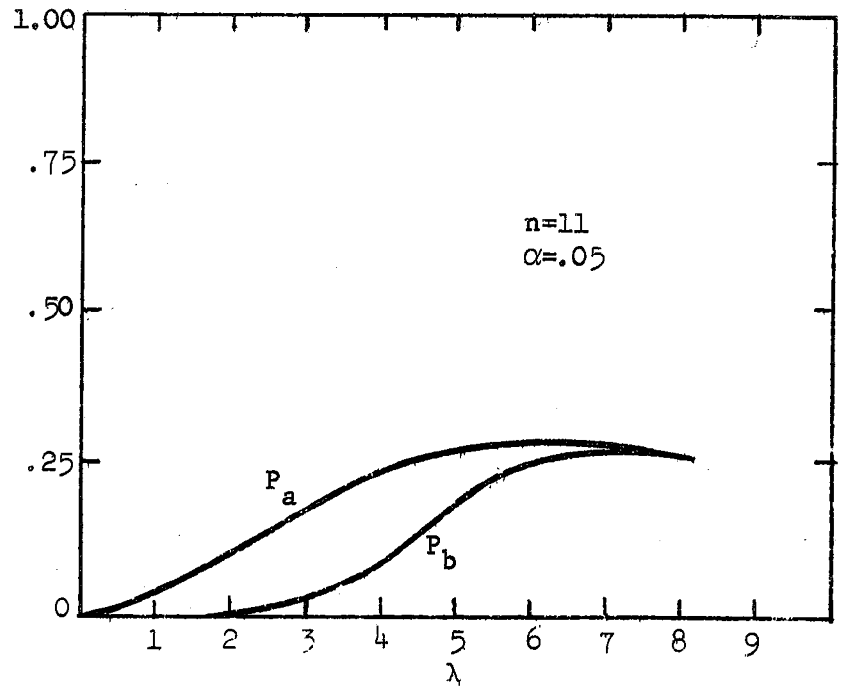
A comparison of these results with those on page 32 indicates that convergence of first stage probabilities is more favorable in testing for one high and one low outlier than in testing for two high outliers. In this statement more favorable convergence means that convergence to 1 occurs for a greater portion of the appropriate quarter plane of λ_1, λ_2. Convergence is still less favorable if we are concerned with two outliers either both positive or both negative. This situation would require use of the rejection constants for $\frac{|x-\bar{x}|_{(n)}}{S}$, which are larger than for $(x_{(n)}-\bar{x})/S$, but for the first stage probabilities to converge to 1 we would have to have overlapping hyperbolas in the first and third quadrants.

Figure 2.8 shows the performance of this test against the alternative of one high and one low outlier given in (2.10). For n = 11, the convergence to 0 is slower than in the one-sided case. This might be expected, since $\gamma_o = .803$ for the one sided-case and $\gamma_1 = -.897$ for the two-sided case. For n = 21, the asymptotes favor the one-sided case. Nevertheless the two-sided performance seems to be more favorable for the larger values of λ, indicating the difference in correlations of residuals is important.

2.2.7 External Studentization

One way of escaping part of the masking effect of multiple outliers is to studentize with an independent variance estimate. This procedure is obviously superior in some circumstances, since the test based on combined internal and external studentization may have decreasing probability of detecting even one outlier as λ increases. It may happen that external studentization is preferable in other cases also, as shown by the examples of Figure 2.9.

With results for additional cases, one could perhaps determine for each sample size the minimum ν for which external studentization would be in some sense superior. The incentive to do so is somewhat limited, since if one knew the alternative to specify exactly two outliers there are better tests, as discussed in Chapter 3.



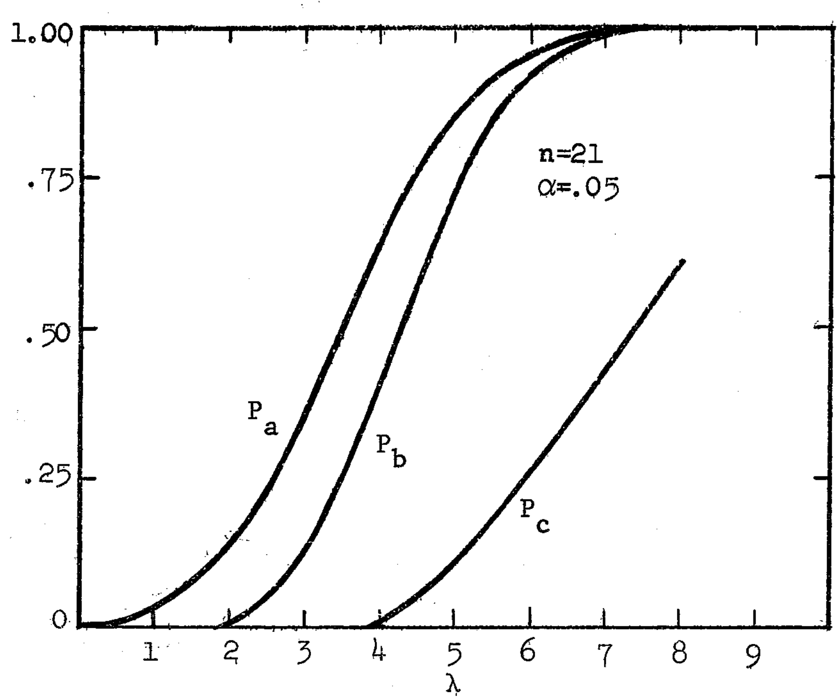


Figure 2.8 Performances P_a, P_b, P_c of sequential maximum residual test of section 2.2.6 when $\alpha = .05$ and $x_1 \sim N(\mu+\lambda, \sigma^2)$, $x_2 \sim N(\mu-\lambda, \sigma^2)$, $x_3, \ldots, x_n \sim N(\mu, \sigma^2)$

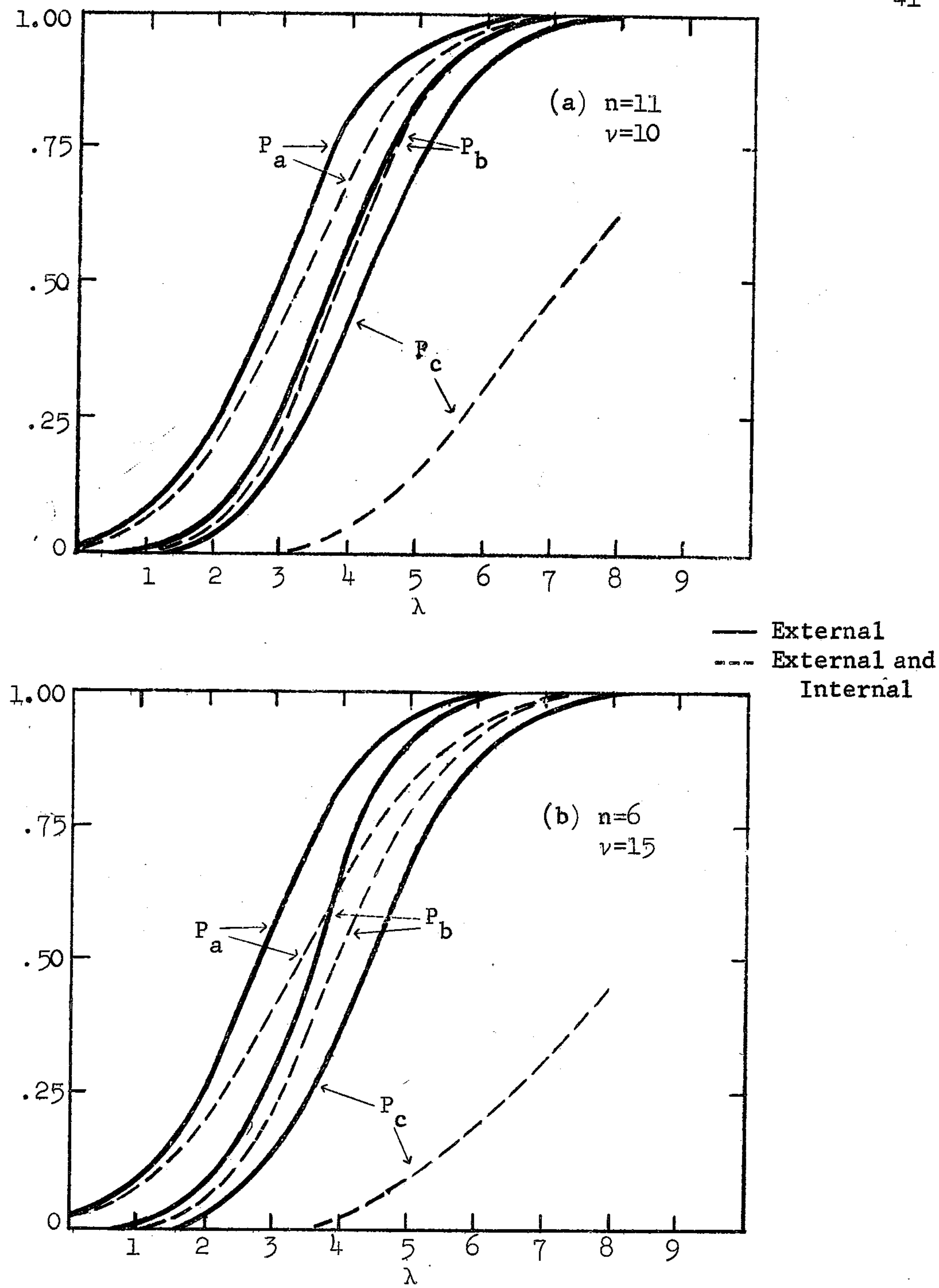


Figure 2.9 Performances P_a, P_b, P_c of externally studentized sequential maximum residual test of section 2.2.7 when $\alpha = .05$ and $x_1, x_2 \sim N(\mu+\lambda, \sigma^2)$, $x_3, \ldots, x_n \sim N(\mu, \sigma^2)$, showing the performances of the test of section 2.2.1 as dashed lines for comparison

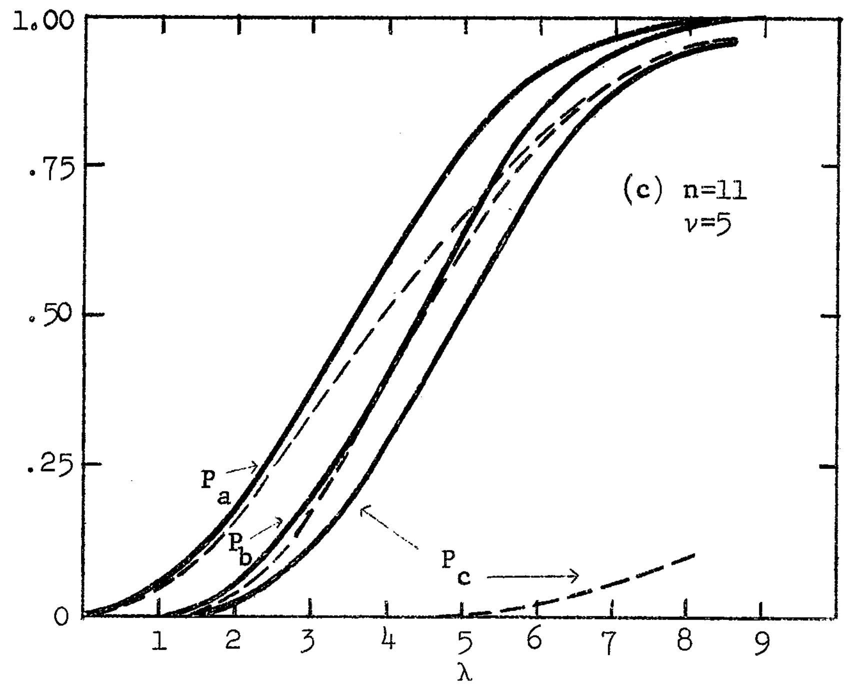


Figure 2.9 (continued)

The numerical calculation of P_a, P_b, and P_c for external studentization were obtained by numerical integration. For this test procedure the first and second stage test statistics $(x_1-\bar{x})/s_\nu$ and $(x_2-\bar{x}_1)/s_\nu$ are not independent. For fixed s_ν, however, P_a, P_b, and P_c are easily determined, and can be integrated with respect to the pdf of s_ν.

3. MURPHY'S TEST FOR TWO OUTLIERS

3.1 Introduction

Suppose that in a sample of n normally distributed observations, K observations are distributed $N(\mu+\lambda, \sigma^2)$, $\lambda \geq 0$, and n-K observations are distributed $N(\mu, \sigma^2)$. Among tests invariant under change of scale and location, Murphy (1951) showed that the most powerful test of the hypothesis $\lambda = 0$ against the alternative $\lambda > 0$ is: declare the K largest observations outliers if

$$(3.1) \qquad \frac{x_{(n)} + x_{(n-1)} + \dots + x_{(n-K+1)} - K\bar{x}}{\sigma} > b_{n,\alpha}$$

when σ is known, and

$$(3.2) \qquad \frac{x_{(n)} + x_{(n-1)} + \dots + x_{(n-K+1)} - K\bar{x}}{s} > c_{n,\alpha}$$

if σ is unknown. In this chapter rejection constants are presented for some tests with K = 2. When this is not feasible bounds and approximations are given. Finally, simpler proofs of some of Murphy's results are included.

3.2 Known Variance

3.2.1 Rejection Constants

For the case of known variance we assume $\sigma^2 = 1$. The first problem is the development of suitable rejection constants, $b_{n,\alpha}$, such that

$$(3.3) \qquad \Pr\{ x_{(n)} + x_{(n-1)} - 2\bar{x} > b_{n,\alpha} \} = \alpha$$

when in fact there are no outliers. Using an approach similar to that of Grubbs (1950) and Nair (1948) in obtaining the distribution of the largest residual, a method of obtaining α for given b will be developed. Then b corresponding to a prescribed α can be obtained by inverse interpolation.

Following Grubbs and using his notation in this section, the density of the ordered observations $x_1 < x_2 < \ldots < x_n$ is

$$f(x_1, x_2, \ldots, x_n) = \frac{n!}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2} .$$

Making the Helmert transformation

$$2.1\ \eta_2 = -x_1 + x_2$$

$$3.2\ \eta_3 = -x_1 - x_2 + 2x_3$$

$$\vdots$$

$$n(n-1)\ \eta_n = -x_1 - x_2 - \ldots - x_{n-1} + (n-1)x_n$$

$$n\ \eta_{n+1} = x_1 + x_2 + \ldots + x_n$$

we have

$$f(\eta_2, \ldots, \eta_n, \eta_{n+1}) = \frac{n!}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=2}^{n+1} \eta_i^2} \qquad 0 \le \eta_2$$

$$\eta_{r-1} \le \sqrt{\frac{r}{r-2}}\ \eta_r,$$

$$r = 3, \ldots, n$$

$$-\infty < \eta_{n+1} < \infty$$

and

$$f(\eta_2,\ldots,\eta_n) = \frac{n!}{(2\pi)^{(n-1)/2}} \; e^{-\frac{1}{2}\sum_{i=2}^{n} \eta_i^2} \quad .$$

Making the further transformation

$$u_r = \frac{\sqrt{r(r-1)}}{r} \; \eta_r$$

we have, still following Grubbs and Nair

$$(3.4) \qquad f(u_2,\ldots,u_n) = \frac{n!\,\sqrt{n}}{(2\pi)^{(n-1)/2}} \; e^{-\frac{1}{2}\sum_{2}^{n} \frac{r}{r-1} u_r^2} \qquad \begin{array}{l} 0 \le u_2 \\ (r-1)u_{r-1} \le r u_r \; . \end{array}$$

The u's are closely related to residuals. In fact

$$u_r = x_r - \bar{x}_r$$

where

$$\bar{x}_r = \frac{x_1 + \ldots + x_r}{r} \quad .$$

Grubbs and Nair integrate out $u_2,\ldots,u_{n-1}$ in (3.4). For our purpose we want the joint density of u_{n-1}, u_n, and integrate out only $u_2,\ldots,u_{n-2}$. Using the following functions as defined by Grubbs,

$$F_2(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-x^2} dx$$

$$F_3(u) = \frac{3\sqrt{3}}{\sqrt{2}} \int_0^u \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{3}{2}x^2)} F_2(\tfrac{3}{2}x)\,dx$$

$$\vdots$$

$$F_n(u) = \frac{n\sqrt{n}}{\sqrt{n-1}} \int_0^u \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{n}{n-1}x^2)} F_{n-1}(\tfrac{n}{n-1}x)\,dx$$

we obtain

$$f(u_3,\ldots,u_n) = \frac{n!\sqrt{n}}{(2\pi)^{\frac{n-1}{2}}} \frac{\sqrt{\pi}}{2} e^{-\frac{1}{2}\sum_3^n \frac{r}{r-1}u_r^2} F_2(\tfrac{3}{2}u_3)$$

$$f(u_4,\ldots,u_n) = \frac{n!\sqrt{n}}{(2\pi)^{\frac{n-1}{2}}} \frac{\sqrt{\pi}}{2} \frac{\sqrt{2}\sqrt{2\pi}}{3\sqrt{3}} F_3(\tfrac{4}{3}u_4)$$

etc., until finally

$$f(u_{n-1},u_n) = \frac{n(n-1)}{2\pi}\sqrt{\frac{n}{n-2}}\, F_{n-2}(\tfrac{n-1}{n-2}u_{n-1})\, e^{-\frac{1}{2}[\frac{n-1}{n-2}u_{n-1}^2 + \frac{n}{n-1}u_n^2]}$$

$$0 \leq u_{n-1} \leq \frac{n}{n-1} u_n .$$

The functions $F_n(x)$ have been tabulated by Grubbs for x in intervals of .05 and for $n \leq 25$. This means that numerical integration of $f(u_{n-1},u_n)$ is fairly easy.

Recall that

$$u_n = x_n - \bar{x}$$

$$u_{n-1} = x_{n-1} - \bar{x}_{n-1} = (x_{n-1} - \bar{x}) + \frac{1}{n-1}(x_n - \bar{x}) .$$

We are interested in the region where

$$x_n + x_{n-1} - 2\bar{x} > b$$

or

$$u_{n-1} + \frac{n-2}{n-1} u_n > b \ .$$

To facilitate evaluation, let

$$V_1 = \frac{n-1}{n-2} u_{n-1}$$

$$V_2 = \sqrt{\frac{n}{n-1}} \, u_n \ .$$

Hence

$$f(V_1, V_2) = n \sqrt{\frac{(n-1)(n-2)}{2\pi}} \; F_{n-2}(V_1) \; e^{-\frac{1}{2}\frac{n-2}{n-1} V_1^2} \; \Phi(V_2)$$

where

$$\Phi(V_2) = \frac{1}{\sqrt{2\pi}} \; e^{-\frac{1}{2} V_2^2}$$

$$0 \leq V_1 \leq \frac{\sqrt{n(n-1)}}{n-2} \; V_2 \ .$$

We now want to integrate $f(V_1, V_2)$ over the region

$$\frac{n-2}{n-1} V_1 + \frac{n-2}{n-1} \sqrt{\frac{n-1}{n}} \; V_2 > b$$

<u>i.e.</u> $$V_2 > \sqrt{\frac{n}{n-1}} \left(b \, \frac{n-1}{n-2} - V_1\right)$$

remembering also that

$$V_2 \geq \frac{n-2}{\sqrt{n(n-1)}} V_1 \quad .$$

Hence for any b the probability that (3.3) holds when there are no outliers can be approximated by

$$\alpha(b) = \sum_{i=1}^{N} \Delta n \sqrt{\frac{(n-1)(n-2)}{2\pi}} \; F_{n-2}(V_{1_i}) \; e^{-\frac{1}{2}\frac{n-2}{n-1}V_{1_i}^2} \left[1 - \Phi\left(\max \frac{\sqrt{\frac{n}{n-1}}\left(b\,\frac{n-1}{n-2} - V_{1_i}\right)}{\frac{n-2}{\sqrt{n(n-1)}}\,V_{1_i}}\right)\right]$$

where

$$V_{1_{i+1}} = V_{1_i} + \Delta \qquad i = 1, 2, \ldots, N-1 \quad .$$

Taking trial values of b and computing $\alpha(b)$, b for $\alpha = .05$ and $\alpha = .01$ can then be obtained by interpolation. It was found that approximations to $b_{n,\alpha}$ with uncertainties only in the fourth significant figure can be obtained with $\Delta = .1$, N about 30, and V_{1_i} suitably chosen for the particular n. Results are presented in Table 3.1.

3.2.2 Performance

As before we consider the model

$$x_i \sim N(\mu+\lambda, 1) \qquad i = 1, 2$$

$$x_i \sim N(\mu, 1) \qquad i = 3, 4, \ldots, n \; .$$

We use the single measure of performance

$$\Pr\{\, x_1 + x_2 - 2\bar{x} > b_{n,\alpha} \,\}$$

Table 3.1 Rejection constants $b_{n,\alpha}$ for sum of largest two standardized residuals

	α	
n	.05	.01
4	2.388	2.934
5	2.805	3.380
6	3.098	3.693
7	3.326	3.932
8	3.509	4.122
9	3.661	4.279
10	3.791	4.411
11	[illegible]	4.526
12	4.005	4.625
13	4.091	4.714
14	4.170	4.794
15	4.241	4.866
16	4.307	4.931
17	4.368	4.990
18	4.426	5.045
19	4.479	5.097
20	4.526	5.144
21	4.571	5.190
22	4.613	5.232
23	4.654	5.273
24	4.690	5.310
25	4.728	5.346
26	4.764	5.379
27	4.798	5.410

without regard to whether other x's are significant. Since $x_1 + x_2 - 2\bar{x}$ is normally distributed with mean $2(\frac{n-2}{n})\lambda$ and variance $2(\frac{n-2}{n})$, the performance is easily calculated. Some results are shown in Figure 3.1. For a particular α, the curves for two values of n, say n_1 and n_2, intersect at exactly one point. The λ coordinate of this point is

$$\lambda_0 = \frac{b_{n_2\alpha}\sqrt{\frac{n_1-2}{n_1}} - b_{n_1\alpha}\sqrt{\frac{n_2-2}{n_2}}}{2\sqrt{\frac{n_2-2}{n_2}}\sqrt{\frac{n_1-2}{n_1}}\left(\sqrt{\frac{n_2-2}{n_2}} - \sqrt{\frac{n_1-2}{n_1}}\right)} .$$

Suppose $n_2 > n_1$. Then the performance for n_1 will be better than that for n_2 if $\lambda < \lambda_0$. If $\lambda > \lambda_0$ the performance for n_2 is better.

Comparison of the Murphy test and the sequential test of Chapter 2 will be deferred to a later chapter. We note here only that the performance of the Murphy test is not directly comparable with P_b of the previous chapter (P_b being the probability of sequential detection of both outliers) with the same α. This is because the average number of observations rejected in samples with no outliers is different.

3.3 Unknown Variance

3.3.1 The Problem of Rejection Constants

Murphy's test for two outliers with unknown variance is to reject H_o if

$$(3.5) \qquad \frac{x_{(n)} - \bar{x} + x_{(n-1)} - \bar{x}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} > c_{n,\alpha} .$$

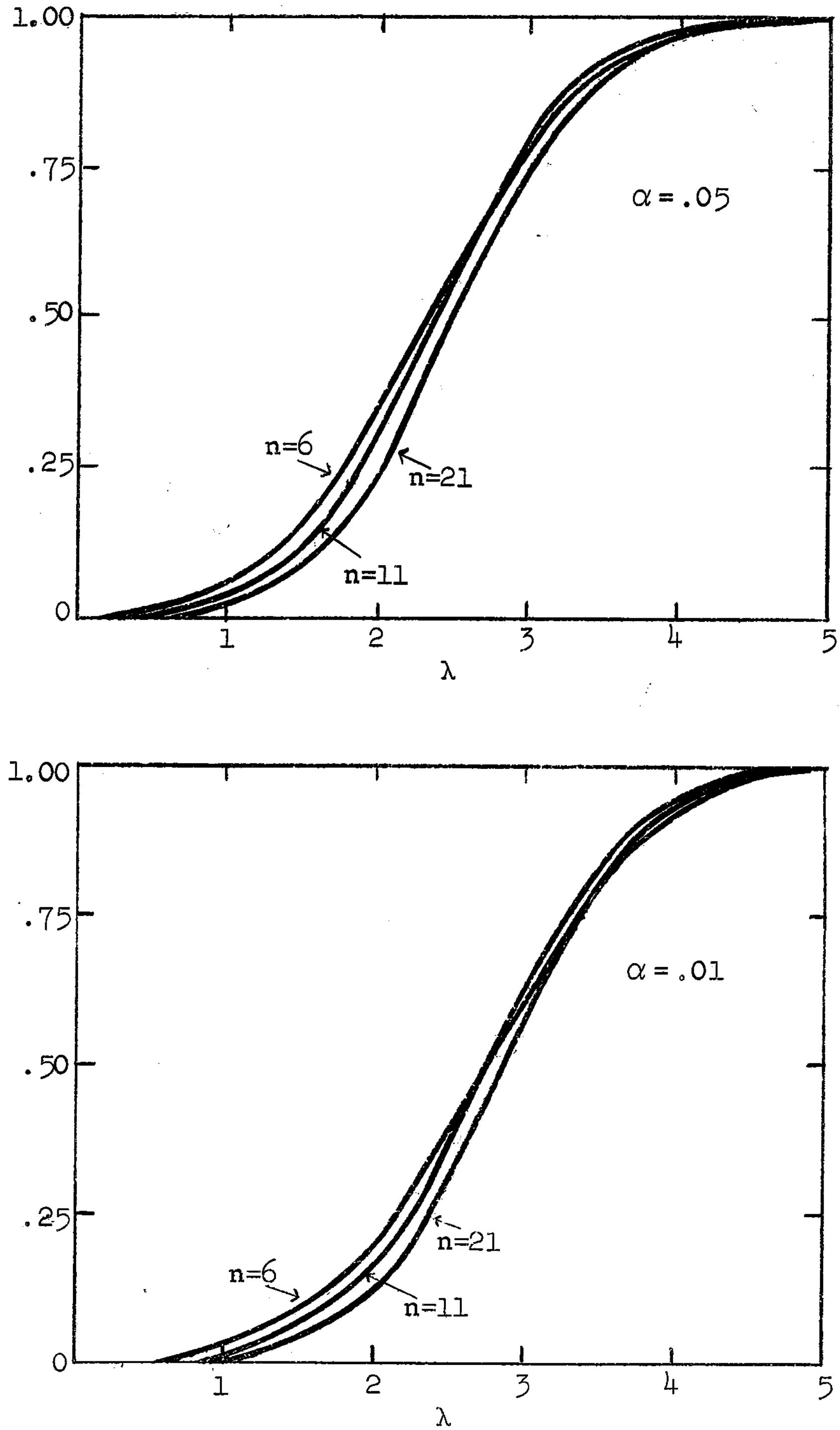


Figure 3.1 Performance of Murphy's test for two outliers when $x_1, x_2 \sim N(\mu+\lambda, 1)$, $x_3, \ldots, x_n \sim N(\mu, 1)$, $\lambda > 0$

To evaluate the performance of the test, or even to apply it, constants $c_{n,\alpha}$ are needed such that (3.5) holds with probability α when there are no outliers. The general solution of this problem is difficult. In some cases, however, it happens that at most one pair of x's can be rejected with a test of the form (3.5), and a simplification results. This occurrence is a generalization of the masking effect mentioned earlier. Previously it was enough to know the upper bound for the next largest studentized residual. Now we need the corresponding bound for the next largest sum of K residuals. From this bound we can obtain immediately the condition under which Murphy's test can reject at most one pair of x's.

3.3.2 Condition for Rejection of at Most One Set of K x's

The following bound for the second largest sum of K studentized residuals from a sample of n was given by Murphy. The derivation included here is simpler than that of Murphy, and uses an approach similar to that of Pearson and Chandra Sekar (1936) in treating a single ordered studentized residual. There are $\binom{n}{K}$ distinct sets of K residuals in a sample of n observations and corresponding to each a value

$$T_j = \frac{(x_{j_1}-\bar{x}) + (x_{j_2}-\bar{x}) + \ldots + (x_{j_K}-\bar{x})}{\sqrt{\sum_{i=1}^{n} (x_i-\bar{x})^2}}$$

where the indices $j_1, j_2, \ldots, j_K$ are chosen from $1, 2, \ldots, n$. Let $\binom{n}{K} = m$, and $T_{(1)} < T_{(2)} < \ldots < T_{(m)}$ the ordered T's.

Theorem: $$T^2_{(m-1)} \leq \frac{2nK - 2K^2 - n}{2n}$$

Proof: $$T^2_{(m-1)} = \frac{[(x_{(n)}-\bar{x}) + (x_{(n-1)}-\bar{x}) + \ldots + (x_{(n-K+2)}-\bar{x}) + (x_{(n-K)}-\bar{x})]^2}{\sum_{i=1}^{n} (x_i-\bar{x})^2}$$

$$= \frac{[(K-1)(\bar{x}_1-\bar{x}) + (x_{(n-K)}-\bar{x}]^2}{(K-1)(\bar{x}_1-\bar{x})^2+(x_{(n-K+1)}-\bar{x})^2+(x_{(n-K)}-\bar{x})^2+(n-K-1)(\bar{x}_2-\bar{x})^2+(n-4)S_o^2}$$

where $\bar{x}_1 = \sum_{i=n-K+2}^{n} x_{(i)} / K-1$

$$\bar{x}_2 = \sum_{i=1}^{n-K-1} x_{(i)} / n-K-1$$

$$(n-4)S_o^2 = \sum_{i=1}^{n-K-1} (x_{(i)}-\bar{x}_2)^2 + \sum_{i=n-K+2}^{n} (x_{(i)}-\bar{x}_1)^2 .$$

Hence $T^2_{(m-1)}$ attains its maximum when $S_o^2 = 0$ and $x_{(n-K+1)}$ is as small as possible, i.e. $x_{(n-K+1)} = x_{(n-K)}$. The only problem remaining is the spacing between the three groups of x's, i.e.

$$x_{(1)} = x_{(2)} = \ldots = x_{(n-K-1)}$$

$$x_{(n-K)} = x_{(n-K+1)}$$

$$x_{(n-K+2)} = \cdots = x_{(n)} .$$

The problem is unaffected by scale and location, so let

$$x_{(1)} = \cdots = x_{(n-K-1)} = 0$$

$$x_{(n-K)} = x_{(n-K+1)} = 1$$

$$x_{(n-K+2)} = \cdots = x_{(n)} = t .$$

We want to maximize in t the function

$$\frac{[1+(K-1)t-K\bar{x}]^2}{\Sigma(x_i-\bar{x})^2} = \frac{[1+(K-1)t-K\frac{2+(K-1)t}{n}]^2}{(K-1)[t+\frac{2+(K-1)t}{n}]^2+2[1-\frac{1+(K-1)t}{n}]^2+(n-K-1)[\frac{2+(K-1)t}{n}]^2}$$

which after simplification is

$$\frac{[(K-1)(n-K)t+(n-2K)]^2}{t^2(n-K+1)(K-1)n+t[-4n(K-1)]+2n(n-2)} .$$

Differentiating with respect to t, the maximum is found to occur at $t=2$, and consequently

$$T^2_{(m-1)} \leq \frac{2nK-2K^2-n}{2n} .$$

With a little algebraic manipulation this agrees with Murphy's result. For the two outlier case we have

$$(3.6) \qquad T^2_{(m-1)} \leq \frac{3n-8}{2n} .$$

Thus

$$\frac{(x_i-\bar{x})+(x_j-\bar{x})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} > c$$

holds for at most one pair i, j provided

$$(3.7) \qquad c \geq \sqrt{\frac{3n-8}{2n}} .$$

We then have by symmetry under H_o

$$(3.8) \qquad \alpha(x) = \Pr \left\{ \frac{(x_{(n)}-\bar{x}) + (x_{(n-1)}-\bar{x})}{\sqrt{\sum_{i=1}^{n} (x_i-\bar{x})^2}} > c \right\}$$

$$= \binom{n}{2} \Pr \left\{ \frac{x_1-\bar{x} + x_2-\bar{x}}{\sqrt{\sum_{i=1}^{n} (x_i-\bar{x})^2}} > c \right\} .$$

The right hand side of (3.8) can be evaluated as follows:

$$\Pr \left\{ \frac{x_1-\bar{x} + x_2-\bar{x}}{\sqrt{\sum_{i=1}^{n} (x_i-\bar{x})^2}} > c \right\} = \Pr \left\{ \frac{(x_1-\ddot{x} + x_2-\bar{x})^2}{\sum_{i=1}^{n} (x_i-\ddot{x})^2} > c^2; \ \frac{x_1 + x_2}{2} > \bar{x} \right\}$$

$$= \Pr \left\{ \frac{4(\frac{n-2}{n})^2(\bar{x}_1-\bar{x}_2)^2}{2(\frac{n-2}{n})(\bar{x}_1-\bar{x}_2)^2+\chi^2_{n-2}} > c^2; \ \bar{x}_1 > \bar{x} \right\}$$

$$= \Pr \left\{ [4(\tfrac{n-2}{n})^2 - 2(\tfrac{n-2}{n})c^2](\bar{x}_1-\bar{x}_2)^2 > c^2\chi^2_{n-2}; \ \bar{x}_1 > \bar{x} \right\}$$

(3.9)

$$= \Pr \left\{ (\frac{2}{n} - \frac{c^2}{n-2})\frac{[2\,\frac{(n-2)}{n}](\bar{x}_1-\bar{x}_2)^2}{\chi^2_{n-2}/n-2} > c^2; \ \bar{x}_1 > \bar{x} \right\}$$

$$= \Pr \left\{ (\frac{2}{n} - \frac{c^2}{n-2})t^2 > c^2; \ \bar{x}_1 > \bar{x} \right\}$$

$$= \Pr \left\{ t > \sqrt{\frac{c^2}{\frac{2}{n} - \frac{c^2}{n-2}}} \right\}$$

where

$$\bar{x}_1 = \frac{x_1 + x_2}{2}$$

$$\bar{x}_2 = \sum_{i=3}^{n} x_i \;/\; n-2$$

and t has the t distribution with n-2 degrees of freedom. Thus we have that if (3.7) holds, then

$$(3.10) \qquad \alpha(c) = \binom{n}{2} \Pr\left\{ t > \sqrt{\frac{c^2}{\frac{2}{n} - \frac{c^2}{n-2}}} \right\} .$$

The restriction on c which allows at most one pair of x's to be significant is equivalent to a restriction on α. The maximum α which can be evaluated precisely with (3.10) is obtainable from (3.10) with $c = \sqrt{\frac{3n-8}{2n}}$. The results are as follows:

n	Maximum α
3	1.0
4	.901
5	.672
6	.464
7	.304
8	.191
9	.115
10	.069
11	.040
12	.026
13	.013
14	.007
15	.004

These values show that accurate rejection constants can be obtained with (3.8) only for $n \leq 10$ if $\alpha = .05$ and $n \leq 13$ if $\alpha = .01$.

3.3.3 A Bound for Error in α in Other Cases

The bound for error in α derived below may be useful in extending the class of tests slightly beyond those for which $c^2 > \frac{3n-8}{2n}$. In any case, we have by Bonferroni inequalities

$$(3.11) \quad \Pr\{ T_{max} > c \} \leq m \Pr\{ T_{ij} > c \}$$

$$(3.12) \quad \Pr\{ T_{max} > c \} \geq m \Pr\{ T_{ij} > c \} - \sum_{(i,j) \neq (K,\ell)} \Pr\{ T_{ij} > c, T_{K\ell} > c \}$$

where

$$m = \binom{n}{2}$$

$$T_{ij} = \frac{x_i - \bar{x} + x_j - \bar{x}}{\sqrt{\sum_1^n (x_i - \bar{x})^2}}$$

and the last sum in (3.12) extends over all distinct pairs, of which there are $\binom{m}{2}$. We call T_{ij} and $T_{K\ell}$ an overlapping pair of T's if the pairs (i,j) and (K,ℓ) have one element in common. Otherwise they are non-overlapping. Let the number of non-overlapping pairs be N_o and the number of overlapping pairs be N_1. Then

$$N_o = \binom{n}{4} 3 = \frac{n(n-1)(n-2)(n-3)}{8}$$

since we can choose the four indices in $\binom{n}{4}$ ways, and each of these sets can be divided into two pairs in 3 ways. Similarly

$$N_1 = \binom{n}{3} 3 = \frac{n(n-1)(n-2)}{2}$$

since the three indices can be chosen in $\binom{n}{3}$ ways and the common element can then be chosen in three ways. It is easily verified that

$$N_o + N_1 = \binom{m}{2} .$$

The problem now is to compute an upper bound for

$$(3.13) \quad \sum_{(ij)\neq(K\ell)} \Pr\{ T_{ij} > c, \; T_{K\ell} > c \}$$

$$= N_o \Pr\{ T_{12} > c, \; T_{34} > c \}$$

$$+ N_1 \Pr\{ T_{12} > c, \; T_{13} > c \} .$$

The last term is the more important. It will be shown later that the term with non-overlapping T's is small in the cases of interest.

Let

$$\bar{x}_1 = \sum_{i=1}^{3} \frac{x_i}{3}$$

$$\bar{x}_2 = \sum_{i=4}^{n} \frac{x_i}{n-3}$$

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$S_1^2 = \sum_{i=1}^{3} (x_i - \bar{x}_1)^2$$

$$S_2^2 = \sum_{i=4}^{n} (x_i - \bar{x}_2)^2$$

then

$$T_{12}^2 = \frac{(x_1 + x_2 - 2\bar{x})^2}{s^2} = \frac{(x_1-\bar{x}_1 + x_2-\bar{x}_1 - 2(\bar{x}-\bar{x}_1))^2}{s_1^2 + \Delta^2 + s_2^2}$$

$$= \frac{[t_1 + t_2 + \frac{2(n-3)}{n}(\bar{x}_1-\bar{x}_2)]^2}{\sum_{i=1}^{3} t_i^2 + \Delta^2 + \chi_{n-4}^2} = \frac{(K\Delta - t_3)^2}{\chi_{n-4}^2 + \Delta^2 + \sum_{i=1}^{3} t_i^2}$$

where

$$\Delta = \sqrt{\frac{3(n-3)}{n}}\,(\bar{x}_1-\bar{x}_2)$$

$$K = 2\sqrt{\frac{n-3}{3n}}$$

$$t_i = x_i-\bar{x}_1 \qquad i = 1,2,3 .$$

For any given χ^2 and Δ, the configuration of t_1, t_2, t_3 that maximizes $\min(T_{12}^2, T_{13}^2)$ is

$$t_2 = t_3 = -Z \qquad \text{(say)}$$

$$t_1 = 2Z .$$

Hence

$$\min(T_{12}^2, T_{13}^2) \leq \frac{(K\Delta + Z)^2}{\chi^2 + \Delta^2 + 6Z^2} .$$

The value of Z which maximizes this function is $\frac{\chi^2 + \Delta^2}{6K\Delta}$. Hence

$$\min(T_{12}^2, T_{13}^2) \leq \frac{(K\Delta + \frac{\chi^2 + \Delta^2}{6K\Delta})^2}{\chi^2 + \Delta^2 + \frac{6(\chi^2 + \Delta^2)^2}{36K^2\Delta^2}} = \frac{K^2\Delta^2}{\chi^2 + \Delta^2} + \frac{1}{6} .$$

Thus for nonzero probability that both T^2_{12}, T^1_{13} exceed c^2 we need

$$c^2 < \frac{K^2\Delta^2}{\chi^2 + \Delta^2} + \frac{1}{6}$$

or

$$\frac{\Delta^2}{\chi^2} > \frac{1}{\dfrac{K^2}{c^2 - \frac{1}{6}} - 1}$$

or

$$(3.14) \qquad t^2_{n-4} > \frac{n-4}{\dfrac{K^2}{c^2 - \frac{1}{6}} - 1} = \frac{n-4}{\dfrac{4(n-3)}{3n(c^2 - \frac{1}{6})} - 1}$$

where t_{n-4} is student's t with n-4 degrees of freedom. An upper limit for the probability that T^2_{12}, T^2_{13} both exceed c^2 can be obtained by finding the probability that (3.14) holds, i.e. the probability that χ^2 and Δ^2 are in such a ratio that simultaneous significance is possible. We have

$$\Pr \{ T^2_{12} > c^2, T^2_{13} > c^2 \} \leq \Pr \{ t^2_{n-4} > \frac{n-4}{\dfrac{4(n-3)}{3n(c^2 - \frac{1}{6})} - 1} \} .$$

Now LHS = $2 \Pr \{ T_{12} > c, T_{13} > c \} + 2 \Pr \{ T_{12} > c, T_{13} < -c \}$, so that

$$(3.15) \qquad \Pr \{ T_{12} > c, T_{13} > c \} \leq \Pr \{ t_{n-4} > \sqrt{\frac{n-4}{\dfrac{4(n-3)}{3n(c^2 - \frac{1}{6})} - 1}} \} .$$

We now consider the simultaneous significance of non-overlapping T's. Since T_{12} and T_{34} are negatively correlated,

$$(3.16)\quad \Pr\{ T_{12} > c,\ T_{34} > c \} < [\Pr\{ T_{12} > c \}]^2, \quad c > 0 .$$

From (3.9), (3.10), (3.12), (3.13), (3.15), and (3.16), we now have the bounds

$$(3.17)\quad \Pr\{ T_{max} > c \} \leq m \Pr\left\{ t_{n-2} > \sqrt{\frac{c^2}{\frac{2}{n} - \frac{c^2}{n-2}}} \right\}$$

$$(3.18)\quad \Pr\{ T_{max} > c \} \geq m \Pr\left\{ t_{n-2} > \sqrt{\frac{c^2}{\frac{2}{n} - \frac{c^2}{n-2}}} \right\}$$

$$- N_o\left[\Pr\left\{ t_{n-2} > \sqrt{\frac{c^2}{\frac{2}{n} - \frac{c^2}{n-2}}} \right\}\right]^2$$

$$- N_1 \Pr\left\{ t_{n-4} > \sqrt{\frac{n-4}{\frac{4(n-3)}{3n(c^2 - \frac{1}{6})} - 1}} \right\} .$$

The second term on the right in (3.18) is relatively small and can frequently be disregarded. For c will usually be established so that $\Pr\{ T_{12} > c \}$ is approximately α/m. If

$$\Pr\{ T_{12} > c \} = \alpha/m$$

the term in question becomes

$$N_o \frac{\alpha^2}{m^2} = \frac{(n-2)(n-3)}{2n(n-1)} \alpha^2 \leq \frac{\alpha^2}{2}$$

which is small enough to be ignored in most cases.

The following table was obtained using (3.17) and (3.18).

n	c	α_{max}	α_{min}
11	.9926	.1375	.02
	1.0489	.0550	.0543
	1.0538	.0500	.0497
15	1.0037	.050	.015
21	1.0064	.0100	.0054

From these results it appears that the simultaneous significance of two T's can safely be ignored only slightly beyond the range of n for which simultaneous significance is impossible. For example the uncertainty in α intended to be .05 is negligible for n = 11 but considerable for n = 15.

3.2.4 Performance

As before we consider the single measure of performance which is the probability of rejection of the two outliers. In this case we compute

$$Pr \left\{ \frac{x_1-\bar{x} + x_2-\bar{x}}{\sqrt{\sum_{i=1}^{n} (x_i-\bar{x})^2}} > c \right\}$$

where

$$x_i \sim N(\mu+\lambda, \sigma^2) \qquad i = 1,2$$

$$x_i \sim N(\mu, \sigma^2) \qquad i = 3,4,\ldots,n \;.$$

For the combinations N, α considered, the uncertainty in α is either nil or negligible although the possibility of simultaneous significance of two T's was disregarded in establishing c. Some results of performance were obtained with the help of noncentral t tables of Locks *et al.* (1963), and are shown in Figure 3.2.

3.4 An Optimum Test for a Predetermined Number of Outliers

Suppose $x_1, x_2, \ldots, x_n$ are independent, normal observations with common variance σ^2 and means μ_i. We want to test the hypothesis H_o: $\mu_i = \mu$ $i = 1, 2, \ldots, n$ against the alternative that the means of $k \leq n$ of the observations have shifted by amounts proportional to known constants. Thus under H_a

$$Ex_1 = \alpha_{i_1}\lambda + \mu$$

$$Ex_2 = \alpha_{i_2}\lambda + \mu$$

$$\vdots$$

$$Ex_n = \alpha_{i_n}\lambda + \mu$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are known and $i_1, \ldots, i_n$ is an unknown permutation of the integers $1, 2, \ldots, n$. Consequently there are $n!$ alternative hypotheses. The α's are not necessarily distinct, and some may be zero, but under the restrictions imposed later they are not all equal, and at least one must be nonzero. Without loss of generality we assume $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n$. We require a procedure for deciding between H_o and one of the $n!$ alternative hypotheses corresponding to the $n!$ permutations. The procedure should have the following properties.

1. The probability of choosing H_o when H_o is true should be $1-\alpha$.

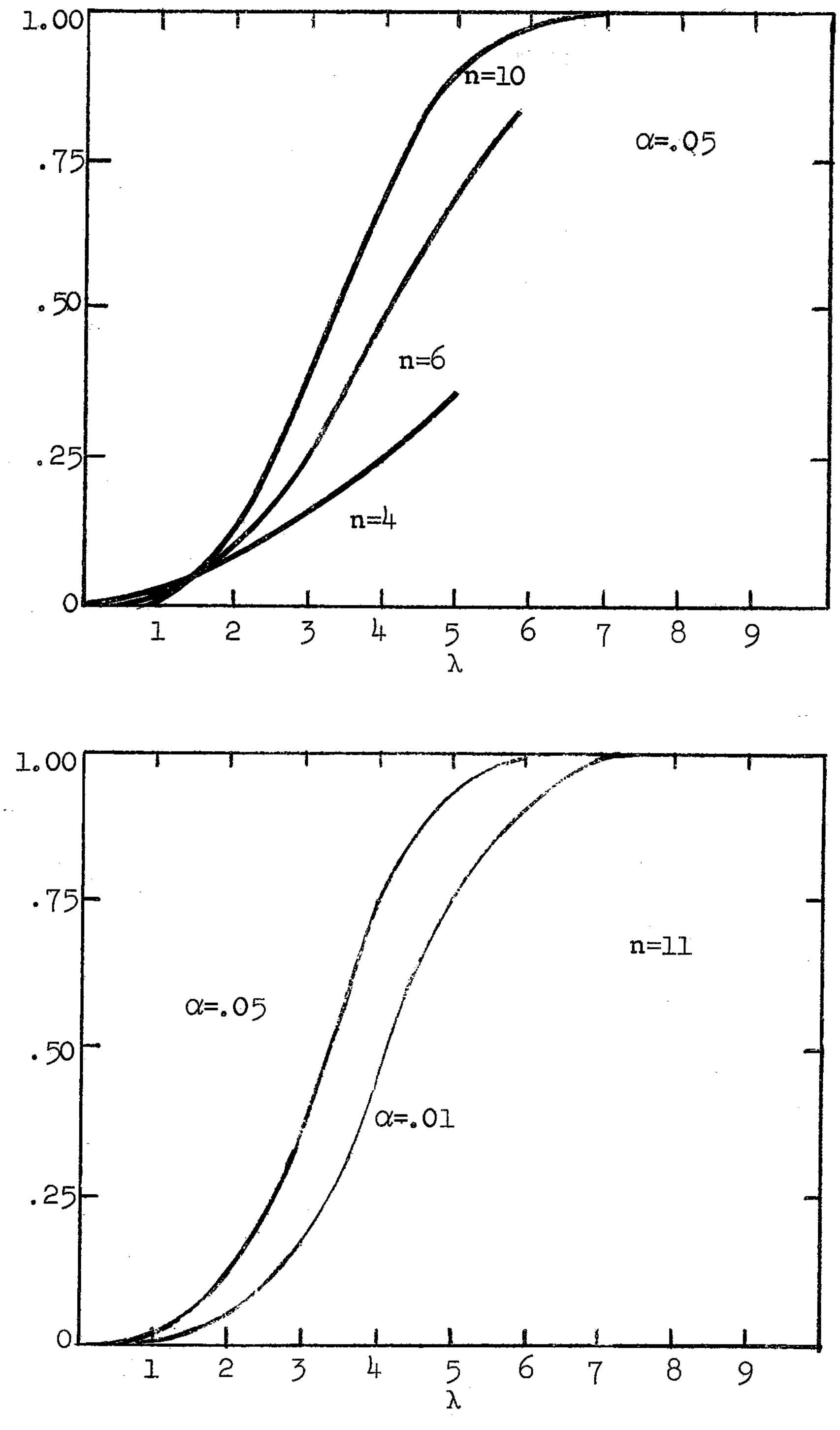


Figure 3.2 Performance of Murphy's test for two outliers when $x_1, x_2 \sim N(\mu+\lambda, \sigma^2)$, $x_3, \ldots, x_n \sim N(\mu, \sigma^2)$, $\lambda > 0$

2. The procedure is unchanged if the same number is added to each observation.

3. The procedure is unchanged if each observation is multiplied by the same positive number.

4. The probability of choosing an alternative if that alternative is true is the same for each of the n! alternatives.

5. The probability of a correct decision is maximum.

Consider first the problem of testing whether a slippage has occurred corresponding to a particular permutation of the α's. For simplicity of notation suppose this permutation takes the subscripts in their natural order. Hence

$$E\, x_i = \mu + \alpha_i \lambda$$

and we want to test

$$H_o: \quad \lambda = 0$$

against

$$H_a: \quad \lambda > 0 \; .$$

In this case

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{(\sigma\sqrt{2\pi})^n} \; e^{-\frac{\Sigma(x_i - \mu - \alpha_i \lambda)^2}{2\sigma^2}}$$

and consequently the statistics Σx_i, $\Sigma\alpha_i x_i$, Σx_i^2 are sufficient for μ, λ, σ^2. Equivalently we can take as sufficient $\bar{x}$, $\Sigma\alpha_i(x_i - \bar{x})$, and s_o^2, where

$$(n-2)s_o^2 = \Sigma(x_i-\bar{x})^2 - \frac{\{\Sigma\alpha_i(x_i-\bar{x})\}^2}{\Sigma(\alpha_i-\bar{\alpha})^2} .$$

Here, s_o^2 is the familiar estimate of residual variance in regression. For invariance under location change the test must depend only on $\Sigma\alpha_i(x_i-\bar{x})$ and s_o^2, and for invariance under scale change as well, it depends only on the ratio $\Sigma\alpha_i(x_i-\bar{x})/s_o$. But this is equivalent to a t statistic with noncentrality parameter $\lambda\Sigma(\alpha_i-\bar{\alpha})^2$, and the test desired rejects H_o if

$$\frac{f_a(t)}{f_o(t)} > c .$$

We use c as a generic constant in the following statements. It has been shown (e.g. Fraser (1957), p. 103) that the preceding statement is equivalent to

$$t > c$$

or

$$\frac{\Sigma\alpha_i(x_i-\bar{x})}{s_o} > c .$$

Using the definition of s_o^2, we can say that the test rejects H_o when

$$\frac{\Sigma\alpha_i(x_i-\bar{x})}{\sqrt{\Sigma(x_i-\bar{x})^2}} > c .$$

Returning to the n+1 decision problem, and applying the generalized Neyman Pearson lemma proved in the appendix, the procedure is:

Choose H_o if $$\frac{\Sigma\alpha_i(x_i-\bar{x})}{\sqrt{\Sigma(x_i-\bar{x})^2}} < c$$

for all permutations of indices of the α's.

Choose H_{a_m} if $$\frac{\Sigma\alpha_i(x_i-\bar{x})}{\sqrt{\Sigma(x_i-\bar{x})^2}} > c$$

for some permutation of the α's, and a_m is the permutation for which the test statistic is maximized.

Ordinarily the α's are not all different, and hence the n! permutations are not all distinct. At times only a restricted set of permutations is relevant because of the nature of the problem. The following examples show some of the more likely applications:

Example 1. The test for K outliers with the same distribution. Suppose we want to choose between the null hypothesis of no outliers and the $\binom{n}{k}$ alternative hypotheses that k observations of unspecified subscript come from a distribution with mean $\mu+\lambda$, $\lambda > 0$. Let $\alpha_1=\alpha_2=\ldots=\alpha_{n-k}=0$, and $\alpha_{n-k+1}=\ldots=\alpha_n=1$. The procedure selects the K largest observations as outliers if

$$\frac{(x_{(n)}-\bar{x}) + (x_{(n-1)}-\bar{x}) + \ldots + (x_{(n-k+1)}-\bar{x})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} > c$$

and otherwise accepts H_o. As special cases we have the test for two outliers considered earlier in this chapter and the usual test for one outlier.

Example 2. The test for one outlier to the right with some observations known to be nonoutliers. Suppose the last r observations contain the outlier, if any. The nonzero α may be taken as 1, and the only permutations of the α's which correspond to distinct alternative hypothesis are the r permutations containing 1 in one of the last r positions and 0's elsewhere. The procedure chooses the largest among the last r observations as an outlier if

$$\max_{i > n-r} \frac{x_i - \bar{x}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} > c .$$

This test is easily generalized to consider the possibility of a number of outliers greater than one.

Example 3. Paulson's result. Suppose the observations are in r groups, with K observations in each group. We want to decide which group, if any, comes from a distribution with mean shifted to the right. Let K of the α's be 1 and all others 0. The only permutations of interest are those in which the nonzero α's fall in the same group. The procedure chooses the group with the largest sample mean if

$$\max_{i} \frac{(\bar{x}_i - \bar{x})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} > c$$

where $\bar{x}_i$ is the sample mean of the i^{th} group.

Example 4. External estimate of variance to be included. Suppose that in addition to the data considered above we have an external estimate of variance s_ν^2, based on ν d.f. The sufficient statistics are now $\bar{x}$, $\Sigma\alpha_i(x_i-\bar{x})$ s_o^2, and s_ν^2. But $\bar{x}$, $\Sigma\alpha_i(x-\bar{x})$, and $s_p^2 = \frac{(n-2)s_o^2 + \nu s_\nu^2}{n-2+\nu}$ are also sufficient. After imposing the requirements of invariance we find the test must depend on

$$\frac{\Sigma\alpha_i(x_i-\bar{x})}{s_p}$$

or equivalently

$$\frac{\Sigma\alpha_i(x_i-\bar{x})}{\Sigma(x_i-\bar{x})^2 + \nu s_\nu^2}$$

which again is a t statistic. The external estimate of variance is essentially the same information as contained in Kudô's third group of observations.

4. GRUBBS' TEST FOR TWO OUTLIERS

4.1 The Test Procedure

The test for two outliers proposed by Grubbs (1950) is to reject the two highest observations if

$$\frac{S^2_{n-1,n}}{S^2} = \frac{\sum_{i=1}^{n-2} (x_{(i)} - \bar{x}_{n-1,n})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} < d_{n,\alpha}$$

where $\bar{x}_{n-1,n}$ is the mean of the observations excluding the two highest. The corresponding statistic can be used to test the two lowest observations. Grubbs derived the distribution of the test statistic and gave a table of rejection constants. The test was proposed on intuitive grounds and no optimum properties are claimed. Since the test is easy to apply, however, and is perhaps widely used, some numerical results for its performance will be given here.

4.2 Performance of Grubbs' Test

As before, we investigate the probability that x_1 and x_2, assumed to be the true outliers, are significant, without regard to whether other observations are significant. The performance index is thus

$$Pr \left\{ \frac{\sum_{i=3}^{n} (x_i - \bar{x}_{12})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} < d_{n,\alpha} \right\}$$

where

$$\bar{x}_{12} = \sum_{i=3}^{n} \frac{x_i}{n-2} \quad .$$

Since

$$\sum_{i=1}^{n} (x_i-\bar{x})^2 = \sum_{i=3}^{n} (x_i-\bar{x}_{12})^2 + \frac{2(n-2)}{n} \left(\bar{x}_{12} - \frac{x_1+x_2}{2}\right)^2 + \frac{1}{2} \; (x-x_2)^2$$

we have as the performance

$$\Pr \left\{ \frac{1}{1 + \dfrac{\frac{2(n-2)}{n} \left(\bar{x}_{12} - \frac{x_1+x_2}{2}\right)^2 + \frac{1}{2} (x_1-x_2)^2}{\sum_{i=3}^{n} (x_i-\bar{x}_{12})^2}} < d_{n,\alpha} \right\}$$

$$= \Pr \left\{ \frac{1}{1 + \dfrac{\chi_2'^2}{\chi_{n-3}^2}} < d_{n,\alpha} \right\}$$

$$(4.1) \qquad = \Pr \left\{ \frac{\chi_2'^2}{\chi_{n-3}^2} > \frac{1}{d_{n,\alpha}} - 1 \right\}$$

$$= \Pr \left\{ F' > \left(\frac{1}{d_{n,\alpha}} - 1\right) \frac{n-3}{2} \right\}$$

where $\chi_2'^2$ is noncentral χ^2 with 2 d.f., χ_{n-3}^2 is ordinary χ^2 with n-3 d.f., and F' is noncentral F with 2 and n-3 d.f. If the two outliers have mean λ, the noncentrality parameter is $2(n-2)/n \; \lambda^2$, say 2K.

Tang (1938) gave the following formula for evaluation of noncentral F probabilities if the degrees of freedom of the denominator χ^2 is even:

$$P_{II} = \Pr\{ F' < F_\alpha \} = e^{-K(1-x)} x^{a+b-1} \sum_{i=1}^{b-1} T_i \tag{4.2}$$

where

$$a = \frac{1}{2}\,[\text{numerator } \chi^2 \text{ degrees of freedom}] = \frac{1}{2}\nu_1$$

$$b = \frac{1}{2}\,[\text{denominator } \chi^2 \text{ degrees of freedom}] = \frac{1}{2}\nu_2$$

$$x = \frac{F_\alpha \frac{\nu_1}{\nu_2}}{1 + F_\alpha \frac{\nu_1}{\nu_2}}$$

$$T_i = \frac{1-x}{ix}\{ (a+b-i+Kx)T_{i-1} + K(1-x)T_{i-2} \}$$

$$T_1 = (a + b - 1 + Kx)\frac{1-x}{x}, \quad T_0 = 1 .$$

The performance of Grubbs' test for n = 11 and n = 21 was evaluated using formula (4.2), with the results shown in Figure 4.1.

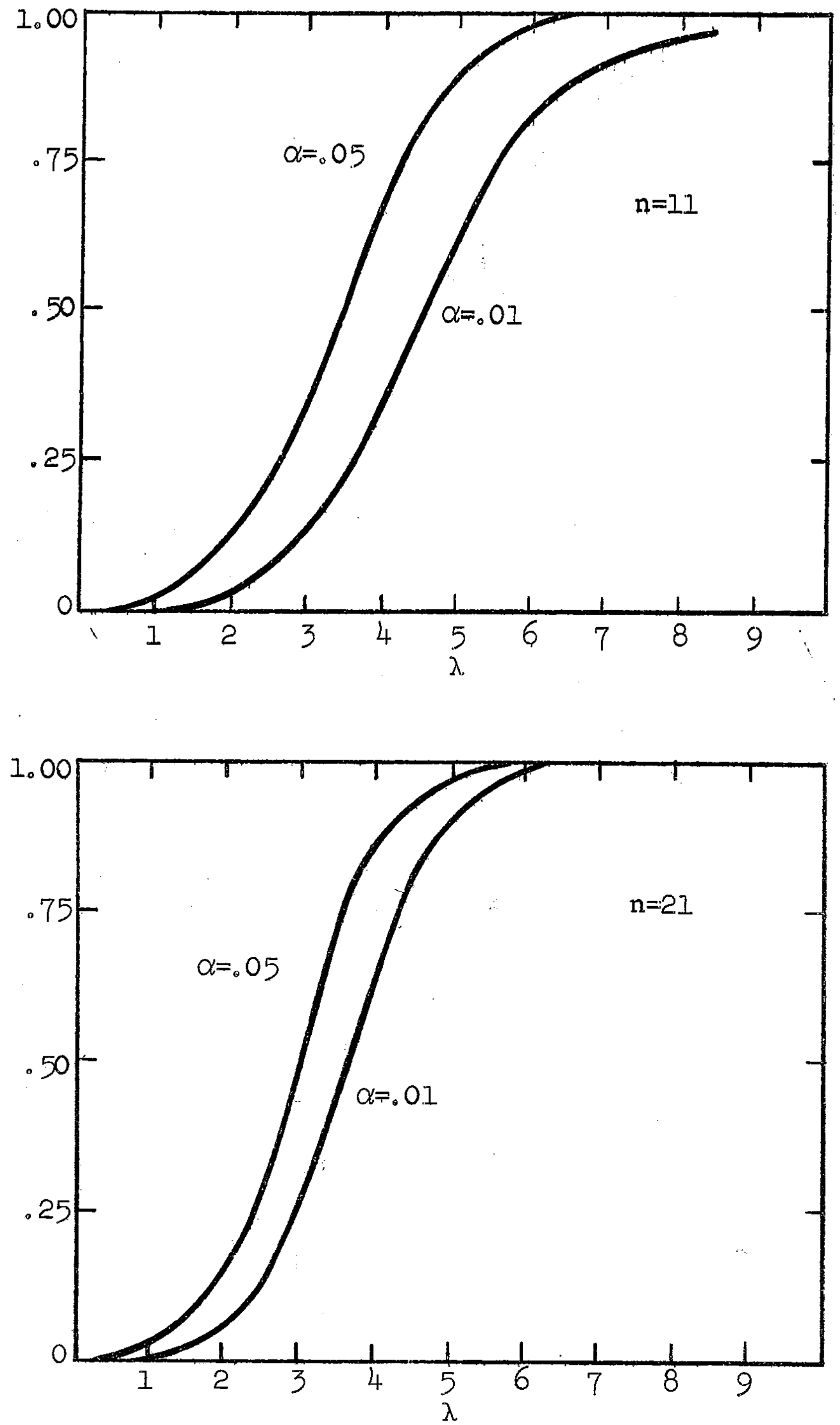


Figure 4.1 Performance of Grubbs' test for two outliers when $x_1, x_2 \sim N(\mu+\lambda, \sigma^2)$, $x_3, \ldots, x_n \sim N(\mu, \sigma^2)$, $\lambda > 0$

5. COMPARISON OF PROCEDURES AND CONCLUSIONS

5.1 Comparison of Murphy's and Grubbs' Tests

The performances with σ^2 unknown of Murphy's (1951) and Grubbs' (1950) tests for two outliers with the same distribution are shown in Figures 5.1 and 5.2 for n=5 and n=11. Murphy's test is superior, as expected from the development of section 3.3 but by a smaller margin for for n=11.

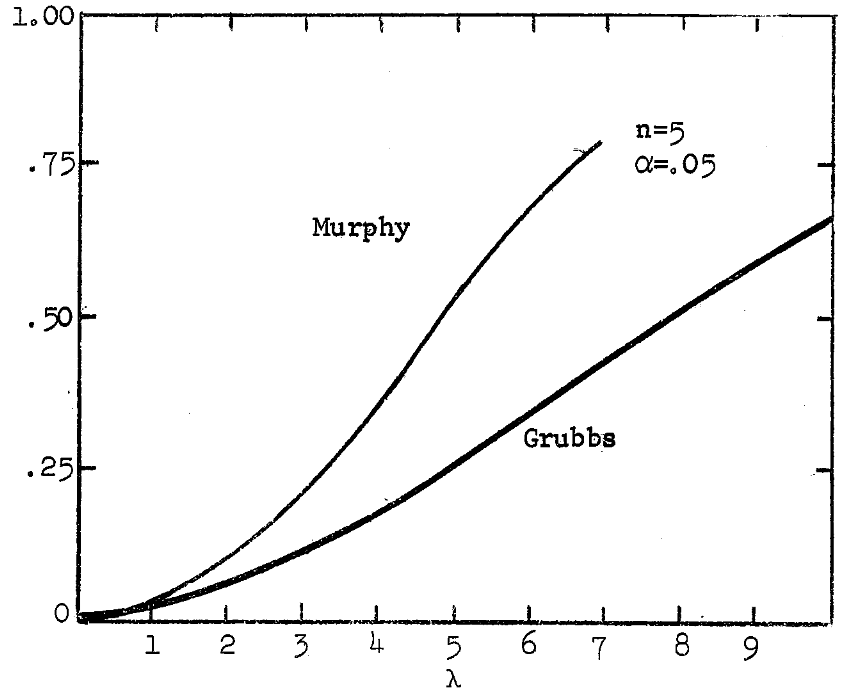
If the outliers are from different distributions, say of means $\mu+\lambda_1$ and $\mu+\lambda_2$, the optimum procedure in the sense of section 3.3 would be based on the weighted sum of two residuals with the weights proportional to λ_1 and λ_2. But since λ_1 and λ_2, or even their ratio, are not likely to be known, it may be of interest to consider the performance of Murphy's test for two outliers from the same distribution when in fact $\lambda_1 \neq \lambda_2$. We recall that the performance of Murphy's test involves the integral of a noncentral t statistic. It follows from (3.9) that the noncentrality parameter is proportional to $\lambda_1+\lambda_2$. Consequently, lines of constant performance in the λ_1, λ_2 plane are of the form $\lambda_1+\lambda_2 = K$.

Performance of Grubbs' test is an integral of the noncentral F distribution of expression (4.1). If the outlier means are unequal, the noncentrality parameter is

$$\left(\frac{\lambda_1+\lambda_2}{2}\right)^2 \frac{2(n-2)}{n} + \frac{(\lambda_1-\lambda_2)^2}{2} = \frac{n-1}{n} (\lambda^2+\lambda_2^2) - \frac{2}{n} \lambda_1\lambda_2$$

and the curves of constant performance are the family

$$(n-1)(\lambda_1^2+\lambda_2^2) - 2\lambda_1\lambda_2 = K .$$



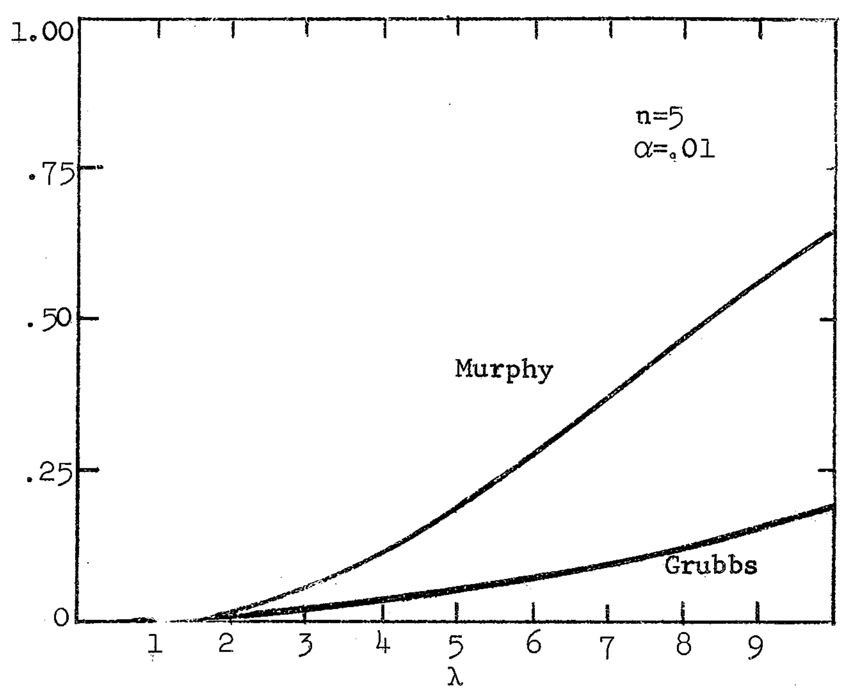


Figure 5.1 Performance of Grubbs' and Murphy's tests when $x_1, x_2 \sim N(\mu+\lambda, \sigma^2)$, $x_3, \ldots, x_n \sim N(\mu, \sigma^2)$, $\lambda > 0$, $n = 5$

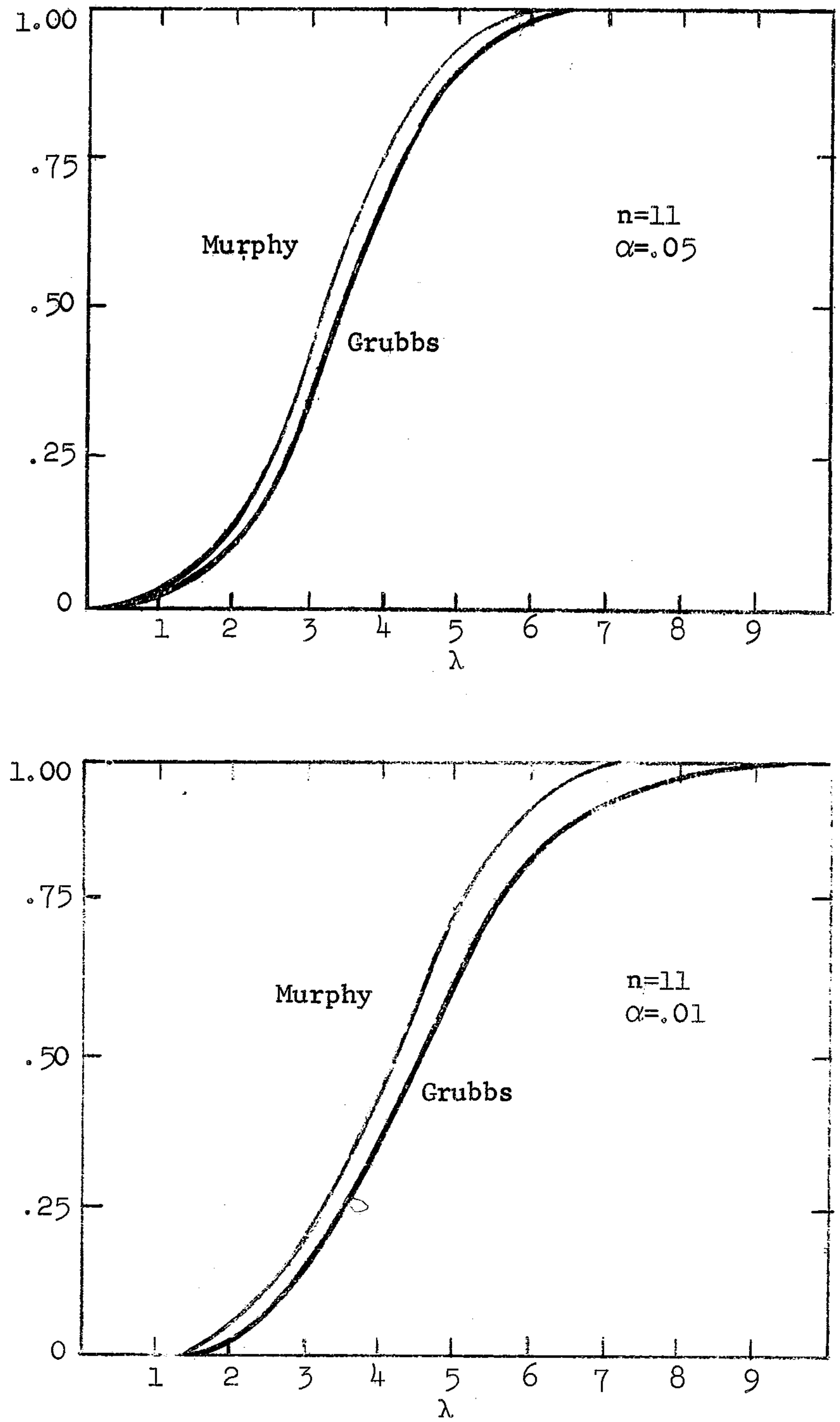


Figure 5.2 Performance of Grubbs' and Murphy's tests when $x_1, x_2 \sim N(\mu+\lambda, \sigma^2)$, $x_3, \ldots, x_n \sim N(\mu, \sigma^2)$, $\lambda > 0$, $n = 11$

Representative lines of constant performance for Murphy's and Grubbs' tests are shown in Figure 5.3. The performances associated with the lines were read graphically, and are accurate to the second place at most. Nevertheless they suffice to show that if the assumption of equal outlier means is not met, the superiority of Murphy's test diminishes rapidly, while Grubbs' test is more robust. If we cannot specify the shift of one outlier relative to the other, but believe both have shifted to the right, we may do better with Grubbs' test.

5.2 Comparison of Three Tests

To compare the sequential test with Murphy's or Grubbs' tests we can consider P_b, the probability of detecting both outliers with the sequential test, along with the single criterion used with the other tests. If $\alpha = .05$ for the sequential procedure, however, the average number of observations rejected per sample with no outliers is $.05 + (.05)^2 = .0525$. Since the average number of rejections with the other tests is 2α, we choose $\alpha = .02625$ to obtain a performance comparable to P_b with the sequential test. The comparison still favors Grubbs' and Murphy's tests somewhat, since the sequential test will detect a single outlier sometimes when it fails to detect both. Also, should it happen that only one outlier exists, the sequential test is preferable.

Grubbs' test is intended for use when σ^2 is unknown. A comparison of the sequential test and Murphy's test for known variance is shown in Figure 5.4. As expected, Murphy's test shows the better performance. The sequential performance is nevertheless reasonably good. A given

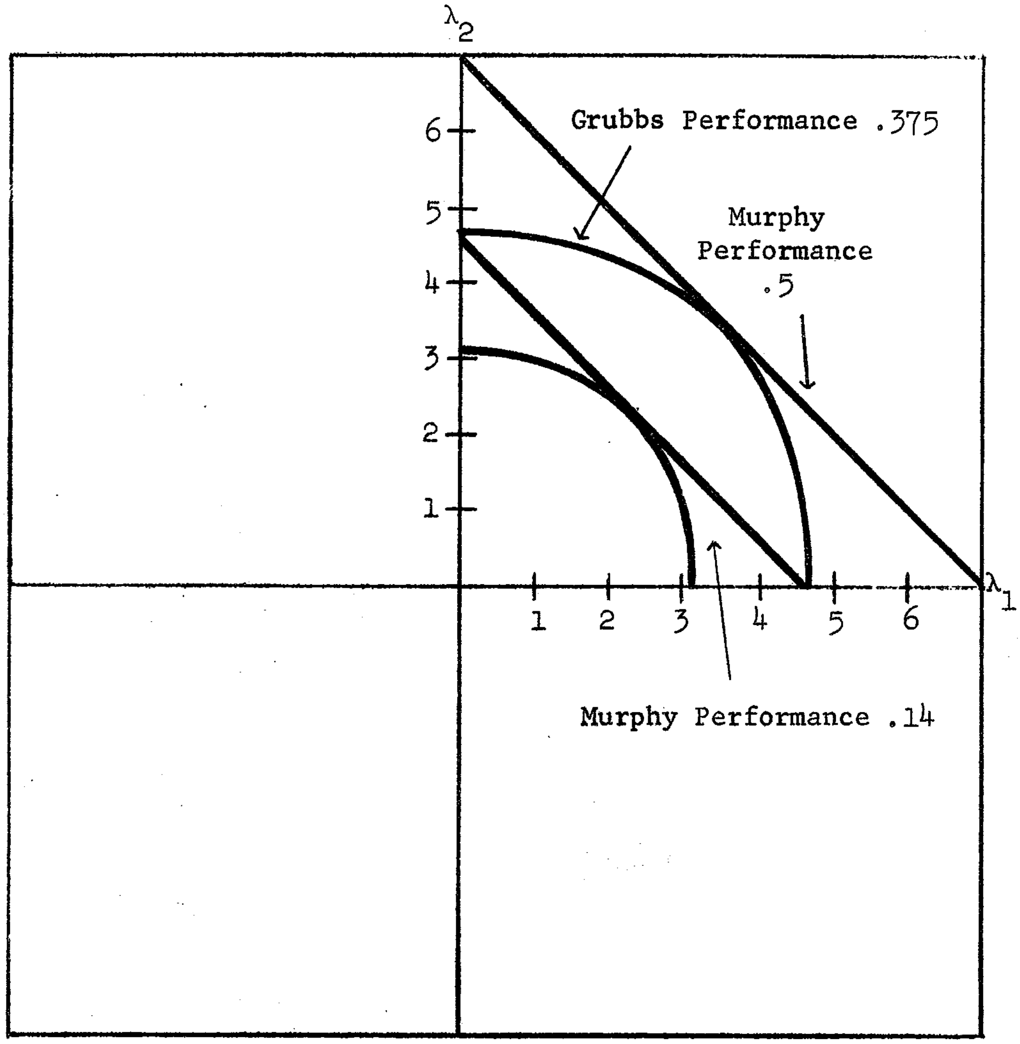


Figure 5.3 Lines of constant performance for Grubbs' and Murphy's tests

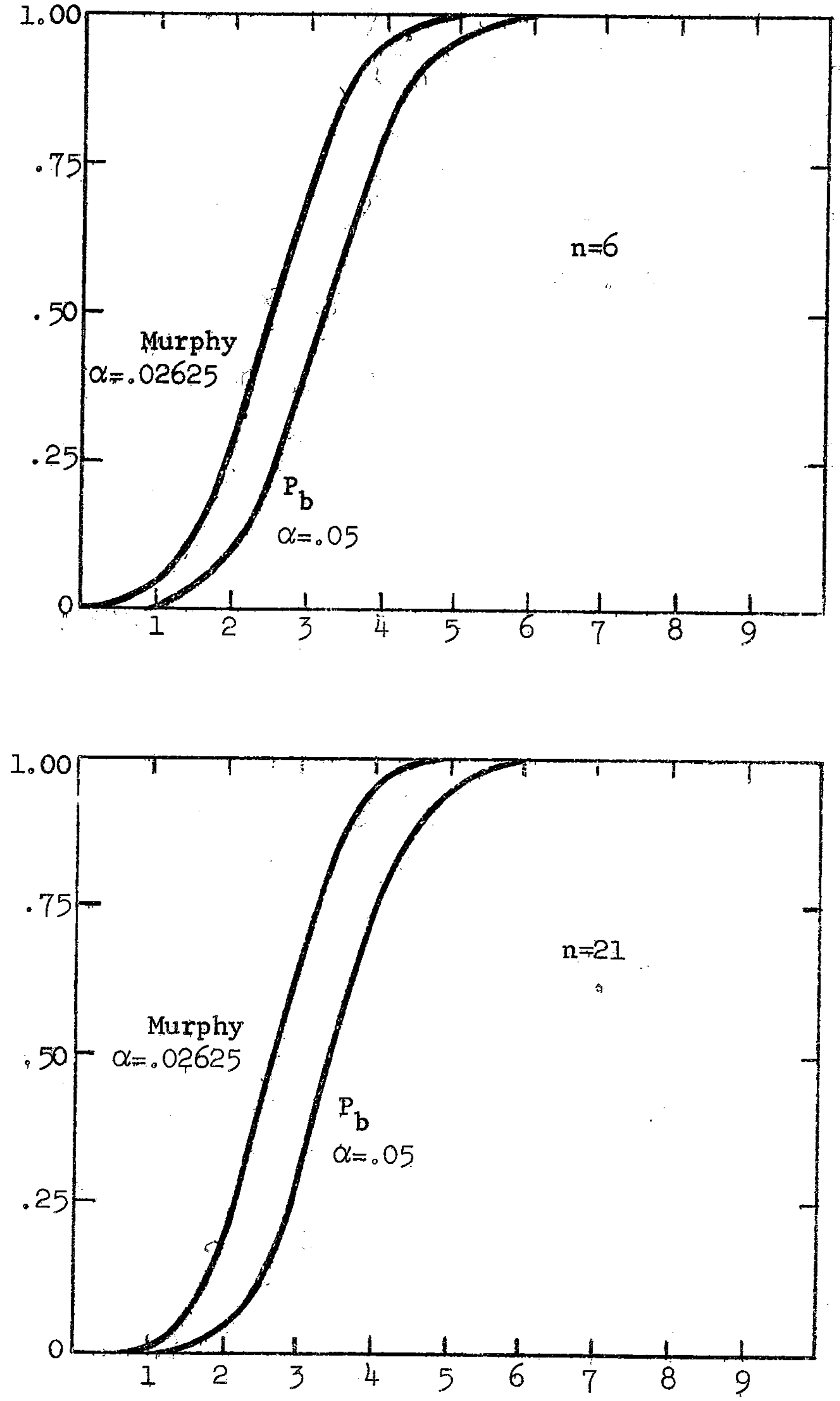


Figure 5.4 Performance of sequential maximum residual and Murphy's tests for two outliers when $x_1, x_2 \sim N(\mu+\lambda, 1)$, $x_3, \ldots, x_n \sim N(\mu, 1)$

performance level can be attained with the sequential procedure with λ about 1σ larger than the λ needed to attain that performance with Murphy's test. This difference is the price paid for not knowing the exact number of potential outliers.

Figure 5.5 shows a comparison of the three tests with unknown variance for n=11 and of Grubbs' and sequential tests with n=21. The rejection constant for Murphy's test for n=21, α=.02625 is not accurately known. The sequential test, compared on this basis, seems to have consistently lower performance than either of the other two tests. The inefficiency of the sequential tests is really serious, however, in those cases for which $P_c \cong 0$. In other cases the sequential procedure is not unreasonable, especially if at most one outlier is expected in the majority of samples to be examined.

5.3 Conclusions

Based on results obtained in this study the following conclusions appear reasonable. The conclusions apply only to outliers displaced in mean, and, except where noted, to problems of at most two outliers.

5.3.1 Samples Expected to Contain at Most One Outlier

If a sample is virtually certain to contain at most one outlier, one of the sequential tests of Chapter 2 is the best procedure for outlier testing. The appropriate form of the test should be used, depending on whether the variance is known or estimated and whether one-sided or two-sided alternatives apply.

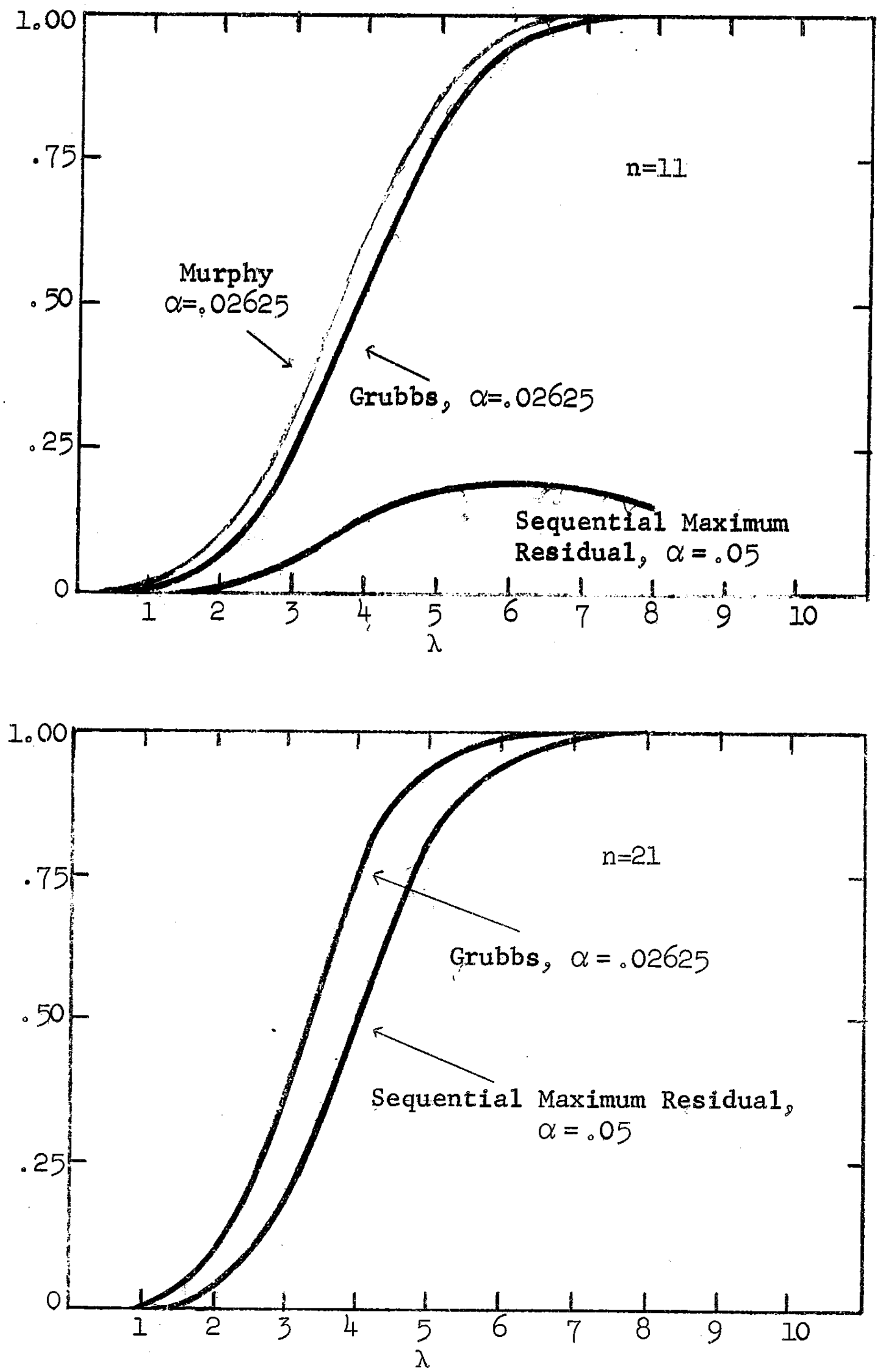


Figure 5.5 Performance of three tests for two outliers when $x_1, x_2 \sim N(\mu+\lambda, \sigma^2)$, $x_3, \ldots, x_n \sim N(\mu, \sigma^2)$

5.3.2 Samples with a Predetermined Number of Potential Outliers

If the number of potential outliers is $K > 1$ (K known), and the choice is between K outliers or no outliers, a test for the particular K should be used. Grubbs' and Murphy's tests are preferable to the sequential test if K=2. If the relative shifts of the outliers are known, as with outliers equal in mean, Murphy's test should be used when rejection constants are available. If the relative shifts cannot be assumed, Grubbs' test is preferable.

5.3.3 Samples with an Unknown Number of Outliers, Known Variance

If the number of potential outliers is unknown and the variance is known, the sequential maximum residual procedure is satisfactory for one or two outliers and $n \geq 6$.

5.3.4 Samples with an Unknown Number of Outliers, Unknown Variance

If the number of potential outliers and the variance are both unknown, a choice of procedures should be made depending on the seriousness of masking by multiple outliers, if present, for the particular n, ν, and α. Under conditions for which $P_c \equiv 0$, *e.g.* $n + \nu < 15$ (approximately) for $\alpha = .05$, the sequential test of maximum residual should be avoided. For somewhat larger $n+\nu$ the sequential procedure is satisfactory.

5.3.5 External Studentization if Multiple Outliers Are Suspected

For cases in which the sequential test is inefficient, as for $n + \nu < 15$, $\alpha = .05$, or marginal, as for $n+\nu$ slightly larger, the test based on external studentization can be preferable to the one based on internal and external studentization even if ν is as low as 5.

6. LIST OF REFERENCES

Anscombe, F. J. 1960. Rejection of outliers. Technometrics 2:123-147.

David, H. A., and A. S. Paulson. 1965. The performance of several tests for outliers. Biometrika 56:429-436.

Dixon, W. J. 1950. Analysis of extreme values. Annals of Mathematical Statistics 21:488-506.

Ferguson, T. S. 1961. Rules for rejection of outliers. Revue Inst. de Stat. 3:29-42.

Fraser, D. A. S. 1957. Nonparametric Methods in Statistics. John Wiley and Sons, Inc., New York, N. Y.

Grubbs, F. E. 1950. Sample criteria for testing outlying observations. Annals of Mathematical Statistics 21:27-58.

Halperin, M., S. Greenhouse, J. Cornfield, and J. Zalokar. 1955. Tables of percentage points for the studentized maximum absolute deviate in normal samples. Journal of the American Statistical Association 50:185-195.

Kudō, A. 1956. On the testing of outlying observations. Sankhya 17:67-76.

Locks, M. O., M. J. Alexander, and B. J. Byars. 1963. New tables of the noncentral t distribution. Office of Technical Services, U. S. Department of Commerce, Washington, D. C.

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 27:466-472.

Murphy, R. B. 1951. On tests for outlying observations. Unpublished PhD thesis, Department of Mathematics, Princeton University, Princeton, N. Y. University Microfilm, Ann Arbor, Michigan.

Nair, K. R. 1948. The distribution of the extreme deviate from the sample mean and its studentized form. Biometrika 35:118-144.

National Bureau of Standards. 1959. Tables of the Bivariate Normal Distribution and Related Functions. Applied Mathematics Series 50. Government Printing Office, Washington, D. C.

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 C. Chandra Sekar. 1936. The efficiency of statistical tools and a criterion for the rejection of outlying observations. Biometrika 28:308-320.

Quesenberry, C. P., and H. A. David. 1961. Some tests for outliers. Biometrika 48:379-390.

Tang, P. C. 1938. The power function of the analysis of variance tests with tables and illustrations of their use. Statistical Research Memoirs 2:126-146.

Thompson, W. R. 1935. On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. Annals of Mathematical Statistics 6:214-219.

Wright, T. W. 1884. A Treatise on the Adjustment of Observations by the Method of Least Squares. D. Van Nostrand Company, New York, N. Y.

7. APPENDICES

7.1 Conditions for Which Simultaneous Detection Implies Sequential Detection

To establish conditions under which simultaneous detection of one high and one low outlier implies sequential detection, we use the geometrical representation introduced in section 2.2.3. In Figure 7.1 the situation for n=21, ν=0, α=.05 is shown. Only the hyperbolas for $\chi^2_{.50}$ = 17.338 are drawn. To relate this to the second stage rejection of, e.g., x_2 given that x_1 was rejected at the first stage we have

$$(x_2-\bar{x}_1)^2 > (W_\alpha^{(n-1,\nu)})^2 s_1^2 = (W_\alpha^{(n-1,\nu)})^2 [\frac{n-1}{n-2} (x_2-\bar{x}_1)^2 + \chi^2_{n+\nu-3}]$$

which leads to

$$(7.1) \qquad t_2 > \sqrt{\frac{\chi^2}{(\frac{1}{W_\alpha^{(n-1,\nu)}})^2 - \frac{n-1}{n-2}}} - \frac{1}{n-1} t_1$$

$$t_2 < -\sqrt{\frac{\chi^2}{(\frac{1}{W_\alpha^{(n-1,\nu)}})^2 - \frac{n-1}{n-2}}} - \frac{1}{n-1} t_1 \quad .$$

This allows us to show, for the same fixed χ^2, the upper and lower rejection regions for x_1 and x_2 at the second stage. These regions have linear boundaries. The two shaded regions in the figure are the regions in which x_1 can be rejected first as a high outlier, followed by x_2 as a low outlier, and the region in which x_2 is rejected as low followed by x_1 as high. The intersection region is that of simultaneous rejection.

From the figure we see that for $\chi^2_{.5}$, initial significance of t_1 and t_2 implies sequential significance. For an algebraic expression of

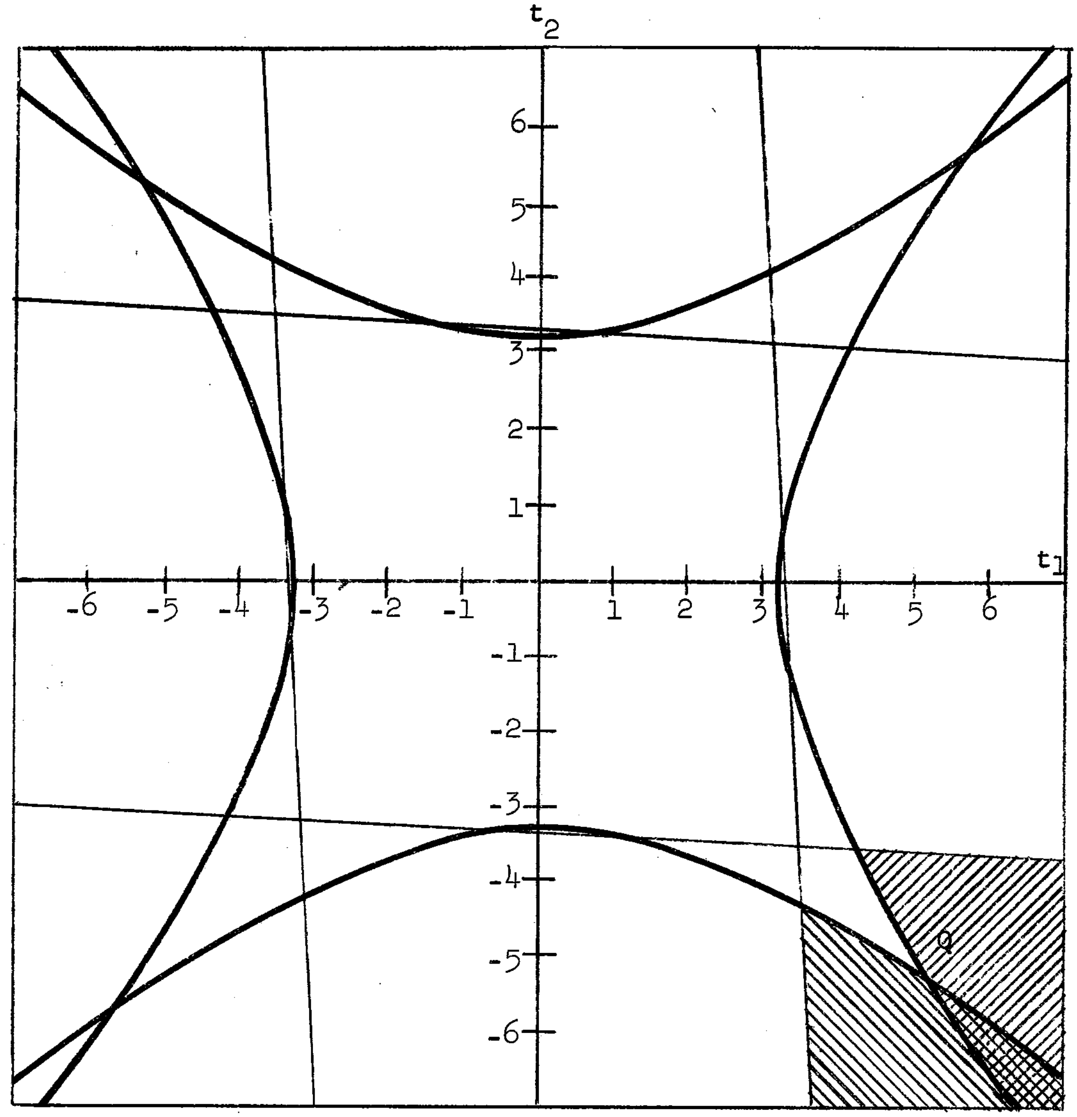


Figure 7.1 First and second stage rejection regions for sequential maximum residual test with σ^2 unknown when n = 21, α = .05, $\chi^2 = \chi^2_{.5}$

a sufficient condition for this, we can require the point of intersection Q to be within the lower second stage x_2 rejection regions (or upper second stage x_1 rejection region). This requirement can be stated

$$(7.2) \qquad \frac{\left(\frac{1}{W_\alpha^{(n,\nu)}}\right)^2 - 2}{\left(\frac{1}{W_\alpha^{(n-1,\nu)}}\right)^2 - \frac{n-1}{n-2}} < \left(\frac{n-2}{n-1}\right)^2 .$$

The requirement does not contain χ^2, and hence holds for all χ^2 if it holds at all.

For n, ν, and α ordinarily of interest, (7.2) can be verified numerically. In the limit as $\nu \longrightarrow \infty$, however, the condition fails. Figure 7.2 shows the first and second stage rejection regions for x_1 and x_2 with known variance, for n=21, α=.05. The first stage rejection regions are

$$|x_i - \bar{x}| = |t_i| > W_{.05}^{(21)} = 2.97 \qquad i = 1,2$$

and are shown as vertical and horizontal dotted lines.

The second stage rejection regions are given by

$$|x_2 - \bar{x}_1| = |t_2 + \frac{1}{n-1} t_1| > W_{.05}^{(20)} = 2.94$$

$$|x_1 - \bar{x}_2| = |t_1 + \frac{1}{n-1} t_2| > W_{.05}^{(20)} = 2.94$$

and are shown as sloping lines. Simultaneous first stage rejection regions are shaded in the first and fourth quadrants, and show the different relation to second stage rejection depending on whether the outliers are on the same side of the mean.

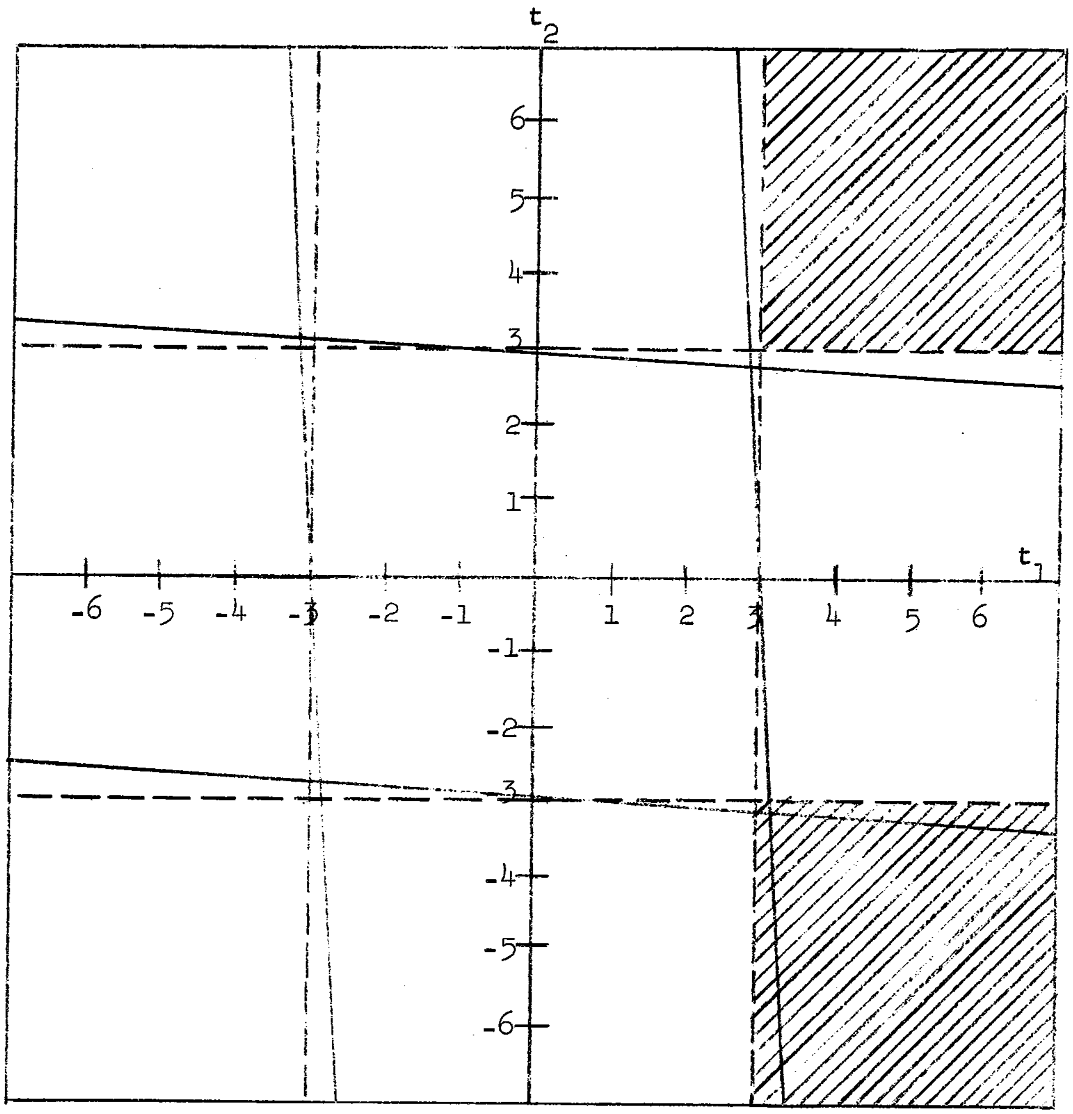


Figure 7.2 First and second stage rejection regions for sequential maximum residual test with σ^2 known when n = 21, α = .05

7.2 A Generalization of the Neyman-Pearson Fundamental Lemma

Suppose we require a procedure for classifying the observations $x_1, x_2, \ldots, x_n$ as having come from one of the densities $f_o, f_1, \ldots, f_m$, such that

(a) Probability of choosing f_o when f_o is correct is $1-\alpha$,

(b) Probability of choosing f_i when f_i is correct is the same for $i = 1, 2, \ldots, m$,

(c) Probability of a correct decision is maximum, subject to (a) and (b).

The procedure is to choose f_o if $cf_o(\underline{x}) \leq f_i(\underline{x})$, $i = 1, 2, \ldots, m$; and choose f_i if $f_i(\underline{x}) > f_j$, $j \neq i$ and $f_i(\underline{x}) > cf_o(\underline{x})$.

Proof: Let the decision functions be $d_i(\underline{x})$, $i = 0, 1, \ldots, m$ with $d_i = 1$ where f_i is selected and $d_i = 0$ otherwise. Consider the integral

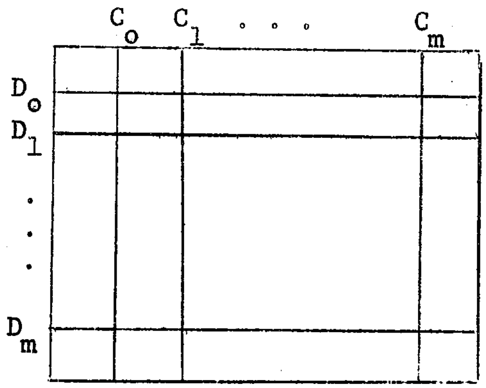
$$\int [c\, d_o(\underline{x}) f_o(\underline{x}) + \sum_{i=1}^{m} d_i(\underline{x}) f_i(\underline{x})] d\underline{x} \quad . \tag{7.3}$$

The required decision procedure is that which maximizes (7.3) subject to (a) and (b). Let $D_o, D_1, \ldots, D_m$ be the partition of the sample space corresponding to the procedure

$$d_o = 1 \quad \text{if} \quad cf_o > f_i \qquad i = 1, 2, \ldots, m$$

$$d_i = 1 \quad \text{if} \quad f_i > f_j, \qquad j \neq i \quad \text{and} \quad f_i > cf_o$$

with c chosen to satisfy (a). Let $C_o, C_1, \ldots, C_m$ be the partition corresponding to any other procedure, say d_i', also satisfying (a) and (b).



Consider the set S_{jK}, for which d selects f_j and d' selects f_K. If j=K, the integral of the form (7.3) over S_{jK} is the same for d and d'. If $j \neq K$ and $j,K > 0$ we have

$$\int_{S_{jK}} [cd_o'(\underline{x}) f_o(\underline{x}) + \sum_{i=1}^{m} d_i'(\underline{x}) f_i(\underline{x})] d\underline{x} = \int_{S_{jK}} f_K(\underline{x}) d\underline{x} \leq \int_{S_{jK}} f_j(\underline{x}) d\underline{x}$$

$$= \int_{S_{jK}} [cd_o(\underline{x}) f_o(\underline{x}) + \sum_{i=1}^{m} d_i(\underline{x}) f_i(\underline{x})] d\underline{x} .$$

Similarly if j=0

$$\int_{S_{oK}} [cd_o'(\underline{x}) f_o(\underline{x}) + \sum_{i=1}^{m} d_i'(\underline{x}) f_i(\underline{x})] d\underline{x} = \int_{S_{oK}} f_K(\underline{x}) d\underline{x} \leq \int_{S_{oK}} cf_o(\underline{x}) d\underline{x}$$

$$= \int_{S_{oK}} [cd_o(\underline{x}) f_o(\underline{x}) + \sum_{i=1}^{m} d_i(\underline{x}) f_i(\underline{x})] d\underline{x}$$

and if K=0

$$\int_{S_{jo}} [cd_o'(\underline{x}) f_o(\underline{x}) + \sum_{i=1}^{m} d_i'(\underline{x}) f_i(\underline{x})] d\underline{x} = \int_{S_{jo}} cf_o(\underline{x}) d\underline{x} \leq \int_{S_{jo}} f_j(\underline{x}) d\underline{x}$$

$$= \int_{S_{jo}} [cd_o(\underline{x}) f_o(\underline{x}) + \sum_{i=1}^{m} d_i(\underline{x}) f_i(\underline{x})] d\underline{x} .$$

Hence the integral (7.3) is smaller with the procedure d' than with d. This completes the proof.

7.3 Tables of Numerical Results

Table 7.1 Performance of sequential maximum residual test of section 2.1.1 when $x_1, x_2 \sim N(\mu+\lambda, 1)$, $x_3, \ldots, x_n \sim N(\mu, 1)$, $\lambda > 0$

n	λ	α = .05 P_a	α = .05 P_b	α = .05 P_c	α = .01 P_a	α = .01 P_b	α = .01 P_c
6	1	.096	.007	.001	.027	.001	.000
	2	.332	.085	.019	.138	.017	.002
	3	.694	.392	.146	.422	.158	.035
	4	.934	.786	.469	.776	.537	.213
	5	.995	.966	.797	.962	.877	.565
	6	1.000	.998	.954	.998	.986	.855
	7		1.000	.993	1.000	.999	.971
	8			.999		1.000	.996
	9			1.000			1.000
11	1	.080	.003	.001	.024	.000	.000
	2	.346	.065	.028	.156	.014	.004
	3	.753	.370	.222	.512	.162	.075
	4	.966	.789	.626	.872	.573	.374
	5	.997	.968	.911	.989	.903	.770
	6	1.000	.998	.989	1.000	.991	.958
	7		1.000	.999		1.000	.996
	8			1.000			1.000
	9						
21	1	.058	.001	.001	.018	.000	.000
	2	.310	.041	.025	.143	.009	.005
	3	.743	.304	.228	.517	.130	.085
	4	.969	.746	.658	.889	.532	.425
	5	.999	.962	.932	.993	.890	.824
	6	1.000	.998	.994	1.000	.989	.976
	7		1.000	1.000		1.000	.999
	8						1.000
	9						

Table 7.2 Performance of sequential maximum residual test of section 2.1.4 when $x_1 \sim N(\mu+\lambda,1)$, $x_2 \sim N(\mu-\lambda,1)$, $x_3,\ldots,x_n \sim N(\mu,1)$

n	λ	$\alpha = .05$		$\alpha = .01$	
		P_a	P_c	P_a	P_c
6	1	.117	.008	.040	.001
	2	.525	.137	.297	.043
	3	.913	.576	.772	.342
	4	.996	.924	.980	.804
	5	1.0	.996	1.0	.981
	6	1.0	1.0	1.0	.999
11	1	.072	.002	.023	.000
	2	.399	.064	.207	.016
	3	.844	.404	.663	.204
	4	.990	.838	.956	.666
	5	1.0	.984	.999	.948
	6	1.0	.999	1.0	.997
21	1	.043	.001	.014	.000
	2	.292	.029	.142	.007
	3	.754	.271	.554	.121
	4	.976	.733	.920	.535
	5	1.0	.963	.997	.900
	6	1.0	.998	1.0	.992

Table 7.3 **Performance of sequential maximum residual test of section 2.2.1 when $x_1, x_2 \sim N(\mu+\lambda, \sigma^2)$, $x_3, \ldots, x_n \sim N(\mu, \sigma^2)$, $\lambda > 0$, $\alpha = .05$**

n	ν	λ	P_a	P_b	P_c
6	15	1	.071	.004	0
		2	.204	.047	0
		4	.635	.517	.018
		6	.939	.928	.178
		8			.441
11	0	1	.046	.002	0
		2	.111	.008	0
		4	.198	.132	0
		6	.195	.188	0
		8	.156	.155	0
	10	1	.060	.002	0
		2	.206	.036	0
		4	.702	.542	.039
		6	.973	.963	.299
		8	.999	.991	.631
21	0	1	.040	.001	0
		2	.222	.028	0
		4	.631	.449	.020
		6	.953	.939	.217
		8	.998	.993	.544

Table 7.4 **Performance of sequential maximum residual test of section 2.2.1 when** $x_1 \sim N(\mu+\lambda_1,\sigma^2)$, $x_2 \sim N(\mu+\lambda_2,\sigma^2)$, $x_3,\cdots,x_n \sim N(\mu,\sigma^2)$, $\lambda_1 = 2\lambda_2 > 0$, $\alpha = .05$

	λ_1	P_a	P_b	P_c
n=11 ν=0				
	1	.038	.001	0
	2	.120	.005	0
	4	.450	.068	0
	6	.760	.290	0
	8	.915	.609	0
n=11 $\nu=\infty$				
	.668	.032	0	0
	1.77	.183	.009	.003
	2.88	.538	.066	.028
	3.98	.869	.214	.111
	5.08	.985	.431	.245

7.5 **Performance of sequential maximum residual test of section 2.2.6 when $x_1 \sim N(\mu+\lambda,\sigma^2)$, $x_2 \sim N(\mu-\lambda,\sigma^2)$, $x_3,\cdots,x_n \sim N(\mu,\sigma^2)$; $\lambda > 0$, $\alpha = .05$, $\nu = 0$**

n	λ	P_a	P_b	P_c
11	1	.036	0	0
	2	.109	.010	0
	4	.243	.129	0
	6	.280	.258	0
	8	.263	.262	0
21	1	.028	0	0
	2	.137	.013	0
	4	.640	.411	.012
	6	.959	.934	.257
	8	.999	.997	.605

Table 7.6 Performance of sequential maximum residual test of section 2.2.7 with external studentization and with internal and external studentization when $x_1, x_2 \sim N(\mu+\lambda, \sigma^2)$, $x_3, \ldots, x_n \sim N(\mu, \sigma^2)$, $\lambda > 0$, $\alpha = .05$

		P_a		P_b		P_c	
	λ	E[a]	I + E[b]	E[a]	I + E[b]	E[a]	I + E[b]
n=6	2	.267	.204	.076	.051	.022	0
ν=15	4	.836	.635	.664	.528	.366	.018
α=.05	6	.996	.939	.986	.934	.886	.178
	8	1.0	.983	1.0	.983	.996	.441
n=11	2	.231	.206	.054	.041	.029	0
ν=10	4	.810	.702	.571	.558	.417	.039
	6	.987	.973	.969	.966	.913	.299
	8	1.0	.999	1.0	.998	.998	.631
n=11	2	.162	.169	.044	.031	.025	0
ν=5	4	.592	.503	.388	.388	.279	.001
α=.05	6	.912	.792	.834	.785	.724	.018
	8	.989	.999	.980	.949	.951	.099

a/ External.

b/ Internal + External.

Table 7.7 Performance of Murphy's test for two outliers when $x_1, x_2 \sim N(\mu+\lambda, \sigma^2)$, $x_3, \ldots, x_n \sim N(\mu, \sigma^2)$, $\lambda > 0$

Known σ^2

	Performance with α=.05			Performance with α=.01		
λ	n = 6	n = 11	n = 21	n = 6	n=11	n=21
1	.063	.038	.020	.020	.012	.006
2	.354	.311	.240	.187	.164	.121
3	.783	.784	.738	.605	.618	.570
4	.974	.980	.976	.922	.943	.936
5	.999	1.0	1.0	.995	.998	.998

Unknown σ^2

Performance with α = .05

n = 4		n = 6		n = 10	
λ	Perf.	λ	Perf.	λ	Perf.
.866	.027	.968	.023	1.186	.026
1.732	.064	1.937	.092	2.372	.193
2.598	.120	2.905	.237	3.558	.564
3.464	.194	3.873	.442	4.744	.875
4.330	.279	4.841	.654	5.930	.983
5.196	.370	5.810	.820	7.116	.999

Performance with n=11

λ	α=.05	α=.025	α=.01
1.23	.032	.019	.008
2.46	.261	.182	.100
3.69	.701	.583	.409
4.92	.951	.901	.780
6.15	.997	.990	.959
7.38	1.0	1.0	.996

Table 7.8 Performance of Grubbs' test for two outliers when $x_1, x_2 \sim N(\mu+\lambda, \sigma^2)$, $x_3, \ldots, x_n \sim N(\mu, \sigma^2)$, $\lambda > 0$

		Performance		
n	λ	α=.05	α=.025	α=.01
5	1	.029	.014	.006
	2	.060	.030	.012
	4	.178	.091	.036
	8	.514	.299	.129
	12	.798	.545	.264
11	1	.021	.011	.004
	2	.115	.067	.031
	3	.358	.240	.132
	4	.675	.526	.345
	6	.979	.939	.834
21	1	.016	.008	.003
	2	.132	.085	.045
	3	.477	.373	.251
	4	.849	.765	.634
	6	.999	.997	.990