AD781 499

AD-781 499

PROCEDURES FOR DETECTING OUTLYING OBSERVATIONS IN SAMPLES

Frank E. Grubbs

Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

April 1974

DISTRIBUTED BY:

NTIS
National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

Destroy this report when it is no longer needed.
Do not return it to the originator.

Secondary distribution of this report by originating or sponsoring activity is prohibited.

Additional copies of this report may be obtained from the National Technical Information Service, U.S. Department of Commerce, Springfield, Virginia 22151.

ACCESSION for

NTIS White Section ☒

D S Buff Section ☐

UNANNOUNCED ☐

JUSTIFICATION

BY

DISTRIBUTION/AVAILABILITY CODES

Dist.	AVAIL. and/or SPECIAL
A	

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD 781 499

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER BRL REPORT NO. 1713	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Procedures for Detecting Outlying Observations in Samples		5. TYPE OF REPORT & PERIOD COVERED Final 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Frank E. Grubbs		8. CONTRACT OR GRANT NUMBER(s) 1T161102A14B
9. PERFORMING ORGANIZATION NAME AND ADDRESS USA Ballistic Research Laboratories Aberdeen Proving Ground, Md. 21005		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS US Army Materiel Command 5001 Eisenhower Avenue Alexandria, VA 22304		12. REPORT DATE April 1974 13. NUMBER OF PAGES ~~56~~ 53
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

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

NATIONAL TECHNICAL INFORMATION SERVICE

18. SUPPLEMENTARY NOTES

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

Detection of Outliers
Discrepant Observations
Data Screening
Significance Tests for Outliers

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

Procedures are given in this report for determining statistically whether the highest observation, or the lowest observation, or the ighest and lowest observations, or the two highest observations, r the two lowest observations, or perhaps more of the observations n the sample may be considered to be outlying observations or iscrepant values. Statistical tests of significance are useful in this connection either in the absence of assignable physical causes

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

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

or to support a practical judgement that some of the experimental observations are aberrant. Both the statistical formulae and illustrative applications of the procedures to practical examples are given, thus representing a rather complete treatment of significance tests for outliers in single univariate samples. This report has been prepared primarily as a useful guide or as an expository and tutorial approach to the problem of detecting outlying observations in much experimental work. We cover only statistical tests of significance in this report and appropriate interpretations which one might draw therefrom.

PROCEDURES FOR DETECTING OUTLYING OBSERVATIONS IN SAMPLES*

SUMMARY

Procedures are given in this report for determining statistically whether the highest observation, or the lowest observation, or the highest and lowest observations, or the two highest observations, or the two lowest observations, or perhaps more of the observations in the sample may be considered to be outlying observations or discrepant values. Statistical tests of significance are useful in this connection either in the absence of assignable physical causes or to support a practical judgement that some of the experimental observations are aberrant. Both the statistical formulae and illustrative applications of the procedures to practical examples are given, thus representing a rather complete treatment of significance tests for outliers in single univariate samples. This report has been prepared primarily as a useful guide or as an expository and tutorial approach to the problem of detecting outlying observations in much experimental work. We cover only statistical tests of significance in this report and appropriate interpretations which one might draw therefrom.

*Substantially the same material of this report is to appear in a *recommended practice* of the American Society for Testing and Materials for dealing with outlying observations. The author is a member of Committee E-11 on Statistical Methods of ASTM and was encouraged by various members of the Society to prepare much of the material covered herein. He gratefully acknowledges their suggestions and advice.

TABLE OF CONTENTS

Preceding page blank

LIST OF TABLES

Preceding page blank

I. INTRODUCTION

This report deals with the problem of outlying observations in samples and how to test the statistical significance of them. An outlying observation, or "outlier", is one that appears to deviate markedly from other members of the sample in which it occurs. In this connection, the following two alternatives are of interest:

(a) An outlying observation may be merely an extreme manifestation of the random variability inherent in the data. If this is true, the values should be retained and processed in the same manner as the other observations in the sample.

(b) On the other hand, an outlying observation may be the result of gross deviation from prescribed experimental procedure or an error in calculating or recording the numerical value. In such cases, it may be desirable to institute an investigation to ascertain the reason for the aberrant value. The observation may even eventually be rejected as a result of the investigation, though not necessarily so. At any rate, in subsequent data analysis the outlier or outliers will be recognized as probably being from a different process or universe than that of the sample values.

(c) It is our purpose here to provide statistical rules that will lead the experimenter almost unerringly to look for causes of outliers when they really exist, and hence to decide whether alternative (a) above is not the more plausible hypothesis to accept, as compared to alternative (b), in order that the most appropriate action in further data analysis may be taken. The procedures covered herein apply primarily to the simplest kind of experimental data, that is, replicate measurements of some property of a given material, or observations in a supposedly single random sample. Nevertheless, the tests suggested do cover a wide enough range of cases in practice to have broad utility.

II. GENERAL

When the experimenter believes that a gross deviation from prescribed experimental procedure has taken place, the resultant observation should be discarded, whether or not it agrees with the rest of the data and without recourse to statistical tests for outliers. If a reliable correction procedure, for example, for temperature, is available, the observation may sometimes be corrected and retained.

In many cases evidence for deviation from prescribed procedure will consist primarily of the discordant value itself. In such cases it is advisable to adopt a cautious attitude. Use of one of the criteria discussed below will sometimes permit a clear-cut decision to be made. In doubtful cases the experimenter's judgement will have considerable influence. When the experimenter cannot identify abnormal conditions, he should at least report the discordant values and indicate to what extent they have been used in the analysis of the data.

Thus, for purposes of orientation relative to the overall problem of experimentation, our position on the matter of screening samples for outlying observations is precisely the following:

Physical Reason Known or Discovered for Outlier(s):

(a) Reject observation(s).
(b) Correct observation(s) on physical grounds.
(c) Reject it (them) and possibly take additional observation(s).

Physical Reason Unknown - Use Statistical Test:

(a) Reject observation(s).
(b) Correct observation(s) statistically.
(c) Reject it (them) and possibly take additional observation(s).
(d) Employ truncated sample theory for censored observations.

The statistical test may always be used to support a judgment that a physical reason does actually exist for an outlier, or the statistical criterion may be used routinely as a basis to initiate action to find a physical cause.

III. BASIS OF STATISTICAL CRITERIA FOR OUTLIERS

There are a number of criteria for testing outliers. In all of these, the doubtful observation is included in the calculation of the numerical value of a sample criterion (or statistic), which is then compared with a critical value based on the theory of random sampling to determine whether the doubtful observation is to be retained or rejected. The critical value is that value of the sample criterion which would be exceeded by chance with some specified (small) probability on the assumption that all the observations did indeed constitute a random sample from a common system of causes, a single parent population, distribution or universe. The specified small probability is called the "signficance level" or "percentage point" and can be thought of as the risk of erroneously rejecting a good observation. It becomes clear, therefore, that if there exists a real shift or change in the value of an observation that arises from non-random causes (human error, loss of calibration of instrument, change of measuring instrument, or even change of time of measurements, etc.), then the observed value of the sample criterion used would exceed the "critical value" based on random sampling theory. Tables of critical values are usually given for several different significance levels, for example, 5 percent, 1 percent. For statistical tests of outlying observations, it is generally recommended that a low significance level, such as 1 percent, be used and that significance levels greater than 5 percent would not be common practice.

Note 1 - In this report, we will usually illustrate the use of the 5 percent significance level. Proper choice of level in probability depends on the particular problem and just what may be involved, along with the risk that one is willing to take in rejecting a good observation, that is, if the null-hypothesis stating "all observations in the sample come from the same normal population" may be assumed.

It should be pointed out that almost all criteria for outliers are based on an assumed underlying normal (Gaussian) population or distribution. When the data are not normally or approximately normally distributed, the probabilities associated with these tests will be different. Until such time as criteria not sensitive to the normality assumption are developed, the experimenter is cautioned against interpreting the probabilities too literally.

Although our primary interest here is that of detecting outlying observations, we remark that some of the statistical criteria presented may also be used to test the hypothesis of normality or that the random sample taken did come from a normal or Gaussian population. The end result is for all practical purposes the same, that is, we really wish to know whether we ought to proceed as if we have in hand a sample of homogeneous observations.

IV. RECOMMENDED CRITERIA FOR SINGLE SAMPLES

Let the sample of n observations be denoted in order of increasing magnitude by $x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n$. Let x_n be the doubtful value, that is the largest value. The test criterion, T_n, recommended here for a single outlier is as follows:

$$T_n = \frac{x_n - \bar{x}}{s}$$

where

$\bar{x}$ = arithmetic average of all n values, and
s = estimate of the population standard deviation based on the sample data, calculated as follows:

$$s = \left\{\frac{\Sigma(x_i - \bar{x})^2}{n - 1}\right\}^{1/2} = \left\{\frac{n\Sigma x_i^2 - (\Sigma x_i)^2}{n(n - 1)}\right\}^{1/2}$$

If x_1 rather than x_n is the doubtful value, the criterion is as follows:

$$T_1 = \frac{\bar{x} - x_1}{s}$$

The critical values for either case, for the 1 and 5 percent levels of significance, are given in Table 1. Table 1 and the following tables at the end of the report give the "one-sided" significance levels. In many previous treatments of outliers, the tables listed values of significance levels double those in the present report, since it was considered that the experimenter would test either the lowest or the highest observation (or both) for statistical significance, for example. However, to be consistent with actual practice and in an attempt to avoid further misunderstanding single-sided significance levels are tabulated herein so that both viewpoints can be represented.

The hypothesis that we are testing in every case is that all observations in the sample come from the same normal population. Let us adopt, for example, a significance level of 0.05. If we are interested *only* in outliers that occur on the *high side*, we should always use the statistic $T_n = (x_n - \bar{x})/s$ and take as critical value the 0.05 point of Table 1. On the other hand, if we are interested *only* in outliers occurring on the *low side*, we would always use the statistic $T_1 = (\bar{x} - x_1)/s$ and again take as a critical value the 0.05 point of Table 1. Suppose, however, that we are interested in outliers occurring on *either side*, but do not believe that outliers can occur on both sides simultaneously. We might, for example, believe that at some time during the experiment something possibly happened to cause an extraneous variation on the high side or on the low side, but that it was very unlikely that two or more such events could have occurred, one being an extraneous variation on the high side *and* the other an extraneous variation on the low side. With this point of view we should use the statistic $T_n = (x_n - \bar{x})/s$ *or* the statistic $T_1 = (\bar{x} - x_1)/s$ whichever is larger. If in this instance we use the 0.05 point of Table 1 as our critical value, the true significance level would be twice 0.05 or 0.10. If we wish a significance level of 0.05 and not 0.10, we must in this case use as a critical value the 0.025 point of Table 1. Similar considerations apply to the other tests given below.

Example 1. As an illustration of the use of T_n and Table 1, consider the following ten observations on breaking strength (in pounds) of 0.104-in. hard-drawn copper wire: 568, 570, 570, 570, 572, 572, 572, 578, 584, 596. The doubtful observation is the high value, $x_{10} = 596$. Is the value of 596 significantly high? The mean is $\bar{x} = 575.2$ and the estimated standard deviation is $s = 8.70$. We compute

$$T_{10} = \frac{596 - 575.2}{8.70} = 2.39$$

From Table 1, for n = 10, note that a T_{10} as large as 2.39 would occur by chance with probability less than 0.05. In fact, so large a value would occur by chance not much more often than one percent of the time. Thus, the weight of the evidence is against the doubtful value having come from the same population as the others (assuming the population is normally distributed). Investigation of the doubtful value is therefore indicated.

An alternative system, the Dixon criteria, based entirely on ratios of differences between the observations is described in the literature (7) and may be used in cases where it is desirable to avoid calculation of s or where quick judgement is called for. For the Dixon test, the sample criterion or statistic changes with sample size. Table 2 gives the appropriate statistic to calculate and also gives the critical values of the statistic for the 1, 5, and 10 percent levels of significance.

Example 2 - As an illustration of the use of Dixon's test, consider again the observations on breaking strength given in Example 1, and suppose that a large number of such samples had to be screened quickly for outliers and it was judged too time-consuming to compute s. Table 2 indicates use of

$$r_{11} = \frac{x_n - x_{n-1}}{x_n - x_2}$$

Thus, for n = 10,

$$r_{11} = \frac{x_{10} - x_9}{x_{10} - x_2}$$

For the measurements of breaking strength above,

$$r_{11} = \frac{596 - 584}{596 - 570} = 0.462$$

which is a little less than 0,477, the 5 percent critical value for n = 10. Under the Dixon criterion, we should therefore *not* consider this observation as an outlier at the 5 percent level of significance. These results illustrate how borderline cases may be accepted under one test but rejected under another. It should be remembered, however, that the T-statistic discussed above is the best one to use for the single-outlier case, and final statistical judgment should be based on it. See Ferguson (8,9).

Further examination of the sample observations on breaking strength of hand-drawn copper wire indicates that none of the other values need testing.

Note 2 - With experience we may usually just look at the sample values to observe if an outlier is present. However, strictly speaking the statistical test should be applied to all samples to guarantee the significance levels used. Concerning "multiple" tests on a single sample, we comment on this below.

A test equivalent to T_n (or T_1) based on the sample sum of squared deviations from the mean for all the observations and the sum of squared deviations omitting the "outlier" is given by Grubbs (11).

The next type of problem to consider is the case where we have the possibility of two outlying observations, the least and the greatest observation in a sample. (The problem of testing the two highest or the two lowest observations is considered below.) In testing the least and the greatest observations simultaneously as probable outliers in a sample, we use the ratio of sample range to sample standard deviation test of David, Hartley and Pearson (5). The significance levels for this sample criterion are given in Table 3. Alternatively, the largest residuals test of Tietjen and Moore (19) could be used.

Example 3 - There is one rather famous set of observations that a number of writers on the subject of outlying observations have referred to in applying their various tests for "outliers". This classic set consists of a sample of 15 observations of the vertical semi-diameters of Venus made by Lieutenant Herndon in 1846 (2). In the reduction of the observations, Prof. Pierce assumed two unknown quantities and found the following residuals which have been arranged in ascending order of magnitude:

-1.40 in.	-0.24	-0.05	0.18	0.48
-0.44	-0.22	0.06	0.20	0.63
-0.30	-0.13	0.10	0.39	1.01

The deviations - 1.40 and 1.01 appear to be outliers. Here the suspected observations lie at each end of the sample. Much less work has been accomplished for the case of outliers at both ends of the sample than for the case of one or more outliers at only one end of the sample. This is not necessarily because the "one-sided" case occurs more frequently in practice but because "two-sided" tests are much more difficult to deal with. For a high and a low outlier in a single sample, we give two procedures below, the first being a combination of tests, and the second a single test of Tietjen and Moore (19) which may have nearly optimum properties.

For optimum procedures when there is an independent estimate at hand, s^2 of σ^2, see (6).

For the observations on the semi-diameter of Venus given above, all the information on the measurement error is contained in the sample of 15 residuals. In cases like this, where no independent estimate of variance is available (that is, we still have the single sample case), a useful statistic is the ratio of the range of the observations to the sample standard deviation:

$$\frac{w}{s} = \frac{x_n - x_1}{s}$$

where:

$$s^2 = \Sigma \frac{(x_i - \bar{x})^2}{n - 1}$$

If x_n is about as far above the mean, $\bar{x}$, as x_1 is below $\bar{x}$, and if w/s exceeds some chosen critical value, then one would conclude that *both* the doubtful values are outliers. If, however, x_1 and x_n are displaced from the mean by different amounts, some further test would have to be made to decide whether to reject as outlying only the lowest value or only the highest value or both the lowest and highest values.

For this example, the mean of the deviations is $\bar{x} = 0.018$, the sample standard deviation, $s = 0.551$, and

$$w/s = \frac{1.01 - (-1.40)}{0.551} = \frac{2.41}{0.551} = 4.374$$

From Table 3 for $n = 15$, we see that the value of $w/s = 4.374$ falls between the critical values for the 1 and 5 percent levels, so if the test were being run at the 5 percent level of significance, we would conclude that this sample contains one or more outliers. The lowest measurement, -1.40 in., is 1.418 below the sample mean and the highest measurement, 1.01 in., is 0.992 above the mean. Since these extremes are not symmetric about the mean, either both extremes are outliers, or else only -1.40 is an outlier. That -1.40 is an outlier can be verified by use of the T_1 statistic. We have

$$T_1 = (\bar{x} - x_1)/s = \frac{0.018 - (-1.40)}{0.551} = 2.574$$

This value is greater than the critical value for the 5 percent level, 2.409 from Table 1 , so we reject -1.40. Since we have decided that -1.40 should be rejected, we use the remaining 14 observations and test the upper extreme 1.01, either with the criterion

$$T_n = \frac{x_n - \bar{x}}{s}$$

or with Dixon's r_{22}. Omitting -1.40 and renumbering the observations, we compute

$$\bar{x} = \frac{1.67}{14} = 0.119, \quad s = 0.401,$$

and

$$T_{14} = \frac{1.01 - 0.119}{0.401} = 2.22$$

From Table 1, for $n = 14$, we find that a value as large as 2.22 would occur by chance more than 5 percent of the time, so we should retain the value 1.01 in further calculations. We next calculate

$$r_{22} = \frac{x_{14} - x_{12}}{x_{14} - x_3} = \frac{1.01 - 0.48}{1.01 + 0.24} = \frac{0.53}{1.25} = 0.424$$

From Table 2 for n = 14, we see that the 5 percent critical value for r_{22} is 0.546. Since our calculated value (0.424) is less than the critical value, we also retain 1.01 by Dixon's test, and no further values would be tested in this sample.

Note 2 - It should be noted that in a multiplicity of tests of this kind, the final over-all significance level will be somewhat less than that used in the individual tests, as we are offering more than one chance of accepting the sample as one produced by a random operation. It is not our purpose here to cover the theory of multiple tests.

For suspected observations on both the high and low sides in the sample, and to deal with the situation in which some of $k \geq 2$ suspected "outliers" are larger and some smaller than the remaining values in the sample, Tietjen and Moore (19) suggesting the following statistic. Let the sample values be $x_1, x_2, x_3, \ldots x_n$ and compute the sample mean, $\bar{x}$. Then compute the n absolute residuals

$$r_1 = |x_1 - \bar{x}|, \; r_2 = |x_2 - \bar{x}|, \; \ldots \; r_n = |x_n - \bar{x}|$$

Now relabel the original observations $x_1, x_2, \ldots, x_n$ as z's in such a manner that z_i is that x whose r_i is the ith largest absolute residual above. This now means that z_1 is that observation x which is closest to the mean and that z_n is the observation x which is farthest from the mean. The Tietjen-Moore statistic for testing the significance of the k largest residuals is then

$$E_k = \frac{\sum_{i=1}^{n-k} (z_i - \bar{z}_k)^2}{\sum_{i=1}^{n} (z_i - \bar{z})^2}$$

$$\text{where} \quad \bar{z}_k = \sum_{i=1}^{n-k} z_i/(n-k)$$

is the mean of the (n-k) least extreme observations and $\bar{z}$ is the mean of the full sample.

Applying this test to the above data, we find that the total sum of squares of deviations for the entire sample is 4.24964. Omitting -1.40 and 1.01, the suspected two outliers, we find that the sum of squares of deviations for the reduced sample of 13 observations is 1.24089. Then E_2 = 1.24089/4.24964 = .292, and by using Table 12, we find that this observed E_2 is slightly smaller than the 5% critical value of .317, so that the E_2 test would reject both of the observations, -1.40 and 1.01. We would probably take this latter recommendation, since the level of significance for the E_2 test is precisely .05 whereas that for the double application of a test for a single outlier cannot be guaranteed to be smaller than $1 - (.95)^2 = .0975$.

The tables of percentage points E_k were computed by Monte Carlo methods on a high-speed electronic calculator.

We next turn to the case where we may have the two largest or the two smallest observations as probable outliers. Here, we employ a test provided by Grubbs (10), (11) which is based on the ratio of the sample sum of squares when the two doubtful values are omitted to the sample sum of squares when the two doubtful values are included. If simplicity in calculation is the prime requirement, then the Dixon type of test (actually ommitting one observation in the sample) might be used for this case. In illustrating the test procedure, we give the following Examples 4 and 5.

Example 4 - In a comparison of strength of various plastic materials, one characteristic studied was the percentage elongation at break. Before comparison of the average elongation of the several materials, it was desirable to isolate for further study any pieces of a given material which gave very small elongation at breakage compared with the rest of the pieces in the sample. In this example, one might have primary interest only in outliers to the left of the mean for study, since very high readings indicate exceeding plasticity, a desirable characteristic.

Ten measurements of percentage elongation at break made on material No. 23 follow: 3.73, 3.59, 3.94, 4.13, 3.04, 2.22, 3.23, 4.05, 4.11, and 2.02. Arranged in ascending order of magnitude, these measurements are: 2.02, 2.22, 3.04, 3.23, 3.59, 3.73, 3.94, 4.05, 4.11, 4.13. The questionable readings are the two lowest, 2.02 and 2.22. We can test these two low readings simultaneously by using the following criterion of Table 4:

$$\frac{S^2_{1,2}}{S^2}$$

For the above measurements:

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

$$= \frac{10(121.3594) - (34.06)^2}{10} = 5.351,$$

and

$$S^2_{1,2} = \sum_{i=3}^{n} (x_i - \bar{x}_{1,2})^2 = \frac{(n-2)\sum_{i=3}^{n} x_i^2 - \left(\sum_{i=3}^{n} x_i\right)^2}{(n-2)}$$

$$= \frac{8(112.3506) - (29.82)^2}{8}$$

$$= \frac{9.5724}{8} = 1.197$$

[where $\bar{x}_{1,2} = \sum_{i=3}^{n} x_i/(n-2)$]

We find:

$$\frac{S^2_{1,2}}{S^2} = \frac{1.197}{5.351} = 0.224$$

From Table 4 for $n = 10$, the 5 percent significance level for $S^2_{1/2}/S^2$ is 0.2305. Since the calculated value is less than the critical value, we should conclude that both 2.02 and 2.22 are outliers. In a situation such as the one described in this example, where the outliers are to be isolated for further analysis, a significance level as high as 5 percent or perhaps even 10 percent would probably be used in order to get a reasonable size of sample for additional study. The problem may really be one of economics, and we use probability as a sensible basis for action.

Example 5 - The following ranges (horizontal distances in yards from gun muzzle to point of impact of a projectile) were obtained in firings from a weapon at a constant angle of elevation and at the same weight of charge of propellant powder:

4782	4420
4838	4803
4765	4730
4549	4833

It is desired to make a judgment on whether the projectiles exhibit uniformity in ballistic behavior or if some of the ranges are inconsistent with the others. The doubtful values are the two smallest ranges, 4420 and 4549. For testing these two suspected outliers, the statistic $S^2_{1,2}/S^2$ of Table 4 is probably the best to use.

Note 4 - Kudo (15) indicates that if the two outliers are due to a shift in location or level, as compared to the scale s, then the optimum sample criterion for testing should be of the type:
$\min(2\bar{x} - x_i - x_j)/s = (2\bar{x} - x_1 - x_2)/s$ in our Example 5.

The distances arranged in increasing order of magnitude are:

4420	4782
4549	4803
4730	4833
4765	4838

The value of S^2 is 158,592. Omission of the two shortest ranges, 4420 and 4549, and recalculation, gives $S^2_{1,2}$ equal to 8590.8. Thus,

$$\frac{S^2_{1,2}}{S^2} = \frac{8590.8}{158,592} = 0.054$$

which is significant at 0.01 level (See Table 4). It is thus highly unlikely that the two shortest ranges (occurring actually from excessive yaw) could have come from the same population as that represented by the other six ranges. It should be noted that the critical values in Table 4 for the 1 percent level of significance are smaller than those for the 5 percent level. So for this particular test, the calculated value is significant if it is less than the chosen critical value.

By Monte Carlo methods using an electronic calculator, Tietjen and Moore (19) have recently extended the tables of percentage points for the two highest or the two lowest observations to k > 2 highest or lowest sample values. Their results are given in Table 11 where $L_k = \sum_{i=1}^{n-k} (x_i - \bar{x}_k)^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2$ and $\bar{x}_k = \sum_{i=1}^{n-k} x_i/(n-k)$. Note that their L_2 equals our $S^2_{n,n-1}/S^2$ or $S^2_{1,2}/S^2$. The columns headed with an * in Tables 11 and 12 indicate agreement with exact values calculated by Grubbs (1950). These new tables may be used to advantage in many practical problems of interest.

If simplicity in calculation is desired, or if a large number of samples must be examined individually for outliers, the questionable observations may be tested with the application of Dixon's criteria. Disregarding the lowest range, 4420 we test if the next lowest range, 4549 is outlying. With n = 7, we see from Table 2 that r_{10} is the appropriate statistic. Renumbering the ranges as x_1 to x_7, beginning with 4549, we find:

$$r_{10} = \frac{x_2 - x_1}{x_7 - x_1} = \frac{4730 - 4549}{4838 - 4549} = \frac{181}{289} = 0.626$$

which is only a little less than the 1 percent critical value, 0.637, for n = 7. So, if the test is being conducted at any significance level greater than a 1 percent level, we would conclude that 4549 is an outlier. Since the lowest of the original set of ranges, 4420, is even more outlying than the one we have just tested, it can be classified as an outlier without further testing. We note here, however, that this test did not use all of the sample observations.

Rejection of Several Outliers - So far we have discussed procedures for detecting one or two outliers in the same sample, but these techniques are not generally recommended for repeated rejection, since if several outliers are present in the sample the detection of one or two spurious values may be "masked" by the presence of other anomalous observations. Outlying observations occur due to a shift in level (or mean), or a change in scale (that is, change in variance of the observations), or both. Ferguson (8,9) has studied the power of the various rejection rules relative to changes in level or scale. For several outliers and repeated rejection of observations, Ferguson points out that the sample coefficient of skewness,

$$\sqrt{b_1} = \sqrt{n} \sum_{i=1}^{n} (x_i - \bar{x})^3/[(n - 1)^{3/2}S^3]$$

$$= \sqrt{n} \sum_{i=1}^{n} (x_i - \bar{x})^3/[\Sigma(x_i - \bar{x})^2]^{3/2},$$

should be used for "one-sided" tests (change in level of several observations in the same direction), and the sample coefficient of kurtosis,

$$b_2 = n \sum_{i=1}^{n} (x_i - \bar{x})^4/[(n - 1)^2S^4]$$

$$= n \sum_{i=1}^{n} (x_i - \bar{x})^4/[\Sigma(x_i - \bar{x})^2].$$

is recommended for "two-sided" tests (change in level to higher and lower values) and also for changes in scale (variance)(see Note 5). In applying the above tests, the

$\sqrt{b_1}$ or the b_2, or both, are computed and if their observed values exceed those for significance levels given in the following tables, then the observation farthest from the mean is rejected and the same procedure repeated until no further sample values are judged as outliers. (As is well known $\sqrt{b_1}$ and b_2 are also used as tests of normality.)

Note 6 -In the above equations for $\sqrt{b_1}$ and b_2, s is defined as used in this standard:

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

The significance levels in the following tables for sample sizes of 5, 10, 15 and 20(and 25 for b_2) were obtained by Ferguson on an IBM 704 computer using a sampling experiment or "Monte Carlo" procedure. The significance levels for the other sample sizes are from E. S. Pearson, "Table of Percentage Points of $\sqrt{b_1}$ and b_2 in Normal Samples; a Rounding off", *Biometrika*, Vol 52; 1965, pp 282 - 285,

SIGNIFICANCE LEVELS FOR $\sqrt{b_1}$

Significance Level, %										
	5[a]	10[a]	15[a]	20[a]	25	30	35	40	50	60
1	1.34	1.31	1.20	1.11	1.06	0.98	0.92	0.87	0.79	0.72
5	1.05	0.92	0.84	0.79	0.71	0.66	0.62	0.59	0.53	0.49

SIGNIFICANCE LEVELS FOR b_2

Significance Level, %								
	5[a]	10[a]	15[a]	20[a]	25[a]	50	75	100
1	3.11	4.83	5.08	5.23	5.00	4.88	4.59	4.39
5	2.89	3.85	4.07	4.15	4.00	3.99	3.87	3.77

[a]These values were obtained by Ferguson, using a Monte Carlo procedure. For n = 25, Ferguson's Monte Carlo values of b_2 agree with Pearson's computed values.

The $\sqrt{b_1}$ and b_2 statistics have the optimum property of being "locally" best against one-sided and two-sided alternatives, respectively. The $\sqrt{b_1}$ test is good for up to 50 percent spurious observations in the sample for the one-sided case and the b_2 test is optimum in the two-sided alternatives case for up to 21 percent "contaminaton" of sample values. For only one or two outliers the sample statistics of the previous paragraphs are recommended, and Ferguson (8) discusses in detail their optimum properties of *pointing out* one or two outliers.

Instead of the more complicated b_1 and b_2 statistics, one can of course use the Tietjen and Moore Tables 11 and 12 included herewith for sample sizes and percentage points given.

V. RECOMMENDED CRITERION USING INDEPENDENT STANDARD DEVIATION

Suppose that an independent estimate of the standard deviation is available from previous data. This estimate may be from a single sample of previous similar data or may be the result of combining estimates from several such previous sets of data. In any event, each estimate is said to have degrees of freedom equal to one less than the sample size that it is based on. The proper combined estimate is a weighted average of the several values of s^2, the weights being proportional to the respective degrees of freedom. The total degrees of freedom in the combined estimate is then the sum of the individual degrees of freedom. When one uses an independent estimate of the standard deviation, s_ν, the test criterion recommended here for an outlier is as follows:

$$T'_1 = \frac{\bar{x} - x_1}{s_\nu}$$

or:

$$T'_n = \frac{x_n - \bar{x}}{s_\nu}$$

where:

ν = total number of degrees of freedom.

The critical values for T'_1 and T'_n for the 5 percent and 1 percent significance levels are due to David (4) and are given in Table 5. In Table 5 the subscript ν = df indicates the total number of degrees of freedom associated with the independent estimate of standard deviation σ and n indicates the number of observations in the sample under study. We illustrate with an example on interlaboratory testing.

Example 6 - Interlaboratory Testing - In an analysis of interlaboratory test procedures, data representing normalities of sodium hydroxide solutions were determined by twelve different laboratories. In all the standardizations, a 0.1 N sodium hydroxide solution was prepared by the Standard Methods Committee using carbon-dioxide-free distilled water. Potassium acid phthalate (P.A.P.), obtained from the National Bureau of Standards, was used as the test standard.

Test data by the twelve laboratories are given in Table 6. The P.A.P. readings have been coded to simplify the calculations. The variances between the three readings within all laboratories were found to be homogeneous. A one-way classification in the analysis of variance was first analyzed to determine if the variation in laboratory results (averages) was statistically significant. This variation was significant, and indicated a need for action, so tests for outliers were then applied to isolate the particular laboratories whose results gave rise to the significant variation.

Table 7 shows that the variation between laboratories is highly significant. To test if this (very significant) variation is due to one (or perhaps two) laboratories that obtained "outlying" results (that is, perhaps showing non-standard technique), we can test the laboratory averages for outliers. From the analysis of variance, we have an estimate of the variance of an individual reading as 0.008793, based on 24 degrees of freedom. The estimated standard deviation of an individual measurement is

$\sqrt{.008793} = 0.094$ and the estimated standard deviation of the average of three readings is therefore

$0.094 / \sqrt{3} = 0.054$.

Since the estimate of within-laboratory variation is independent of any difference between laboratories, we can use the statistic T'_1 above to test for outliers. An examination of the deviations of the laboratory averages from the grand average indicates that Laboratory 10 obtained an average reading much lower than the grand average, and that Laboratory 12 obtained a high average compared to the over-all average. To first test if Laboratory 10 is an outlier, we compute

$$T' = \frac{1.871 - 0.745}{0.054} = 20.9$$

This value of T' is obviously significant at a very low level of probability ($P \ll 0.01$ - Refer to Table 5 with $n = 12$ and $\nu = 24$ degrees of freedom). We conclude, therefore, that the test methods of Laboratory 10 should be investigated.

Excluding Laboratory 10, we compute a new grand average of 1.973 and test if the results of Laboratory 12 are outlying. We have

$$T' = \frac{2.327 - 1.973}{0.054} = 6.56$$

and this value of T' is significant at $P \ll 0.01$ (Refer to Table 5 with $n = 11$ and $\nu = 24$ degrees of freedom). We conclude that the procedures of Laboratory 12 should also be investigated.

To verify that the remaining laboratories did indeed obtain homogeneous results, we might repeat the analysis of variance omitting Laboratories 10 and 12. The calculations give the results shown in Table 8.

For this analysis, the variation between laboratories is not significant at the 5 percent level and we conclude that all the laboratories except No. 10 and No. 12 exhibit the same capability in testing procedure.

In conclusion, there should be a systematic investigation of test methods for Laboratories No. 10 and No. 12 to determine why their test procedures are apparently different from the other ten laboratories. (In this type of problem, the tables of Greenhouse, Halperin, and Cornfield (13) could also be used for testing outlying laboratory averages.)

VI. RECOMMENDED CRITERIA FOR KNOWN STANDARD DEVIATION

Frequently the population standard deviation σ may be known accurately. In such cases, Table 9 may be used for single outliers and we illustrate with the following example:

Example 7 (σ known) - Passage of the Echo 1 (Balloon) Satellite was recorded on star-plates when it was visible. Photographs were made by means of a camera with shutter automatically timed to obtain a series of points for the Echo path. Since the stars were also photographed at the same times as the Satellite, all the pictures show star-trails and are thus called "star-plates."

The x- and y-coordinate of each point on the Echo path are read from a photograph, using a stereo-comparator. To eliminate bias of the reader, the photograph is placed in one position and the coordinates are read; then the photograph is rotated 180 deg. and the coordinates reread. The average of the two readings is taken as the final reading. Before any further calculations are made, the readings must be "screened" for gross reading or tabulation errors. This is done by examining the difference in the readings taken at the two positions of the photograph.

Table 10 records a sample of six readings made at the two positions and the differences in these readings. On the third reading, the differences are rather large. Has the operator made an error in placing the cross-hair on the point?

For this example, an independent estimate of σ is available since extensive tests on the stereo-comparator have shown that the standard deviation in reader's error is about 4μm. The standard deviation of the difference in two readings is therefore

$$\sqrt{4^2 + 4^2} = \sqrt{32} \text{ or } 5.7 \ \mu\text{m}$$

For the six readings (Table 10), the mean difference in the x-coordinates is $\overline{\Delta x} = 3.5$ and the mean difference in the y-coordinates is $\overline{\Delta y} = 1.8$. For the questionable third reading, we have

$$T'_x = \frac{24 - 3.5}{5.7} = 3.60$$

$$T'_y = \frac{22 - 1.8}{5.7} = 3.54$$

From Table 9 we see that for $n = 6$, values of $T'_{n\infty}$ as large as the calculated values would occur by chance less than 1 percent of the time so that a significant reading error seems to have been made on the third point.

A great number of points are read and automatically tabulated on star-plates. Here we have chosen a very small sample of these points. In actual practice, the tabulations would probably be scanned quickly for very large errors such as tabulator errors; then some rule-of-thumb such as ± 3 standard deviations of reader's error might be used to scan for outliers due to operator error (Note 6). In other words, the data are probably too extensive to allow repeated use of precise tests like those described above (especially for varying sample size), but this example does illustrate the case where σ is assumed known. If gross disagreement is found in the two readings of a coordinate, then the reading could be omitted or reread before further computations are made.

Note 6 - Note that the values of Table 9 vary between about 1.4σ and 3.50σ.

VII. ADDITIONAL COMMENTS

In the above, we have covered only that part of screening samples to detect outliers statistically. However, a large area remains after the decision has been reached that outliers are present in data. Once some of the sample observations are branded as "outliers," then a thorough investigation should be initiated to determine the cause. In particular, one should look for gross errors, personal errors of measurement, errors in calibration, etc. If reasons are found for aberrant observations, then one should act accordingly and perhaps scrutinize also the other observations. Finally, if one reaches the point that some observations are to be discarded or treated in a special manner based solely on statistical judgment, then it must be decided what action should be taken in the further analysis of the data. We do not propose to cover this problem here, since in many cases it will depend greatly on the particular case in hand. However, we do remark that there could be the outright rejection of aberrant observations once and for all on physical grounds (and preferably not on statistical grounds generally) and only the remaining observations would be used in further analyses or in estimation problems. On the other hand, some may want to replace aberrant values with newly taken observations and others may want to "Winsorize" the outliers, that is, replace them with the next closest values in the sample. Also, with outliers in a sample, some may wish to use the median instead of the mean, and so on. Finally, we remark that perhaps a fair or appropriate practice might be that of using truncated sample theory (15) for cases of samples where we have "censored" or rejected some of the observations. We cannot go further into these problems here. For additional reading on outliers, see Refs (1, 3, 4, 14, 16, 17, 18).

Finally, a sample test criterion for non-normality, and hence possibly for outliers, not covered above is the Wilk-Shapiro W statistic for a sample of size n given by

$$W = \{\sum_{i=1}^{[n/2]} a_{n-i+1}(x_{n-i+1} - x_i)\}^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

where

$$x_1 \le x_2 \le x_3 \le \cdots \le x_n, \qquad \bar{x} = \sum_{i=1}^{n} x_i/n,$$

$[n/2]$ is the greatest integer in $n/2$, and the coefficients a_{n-i+1} of the order statistics for $n = 2(1)50$ are given in Shapiro, S. S. and M. B. Wilk, "An Analysis of Variance Test for Non-Normality (Complete Samples)", Biometrika Vol. 52 (1965), pp 591-611, as is also a table of percentage points of W for $n = 3(1)50$.

The Wilk-Shapiro W statistic has been found to be quite sensitive to departures from normality and may compare most favorably with the $\sqrt{b_1}$ and b_2 tests discussed above. In addition, therefore, the W statistic may be used also as a test for outliers, or otherwise general heterogeneity of sample values. Our significance tests given above have been selected and recommended since they generally *"point" out* particular suspected outliers in the sample.

TABLE 1

Table of Critical Values for T(One-Sided Test of T_1 or T_n)
When the Standard Deviation is Calculated from the Same Samples

No. Obs. n	Upper .1% Sig. Level	Upper .5% Sig. Level	Upper 1% Sig. Level	Upper 2.5% Sig. Level	Upper 5% Sig. Level	Upper 10% Sig. Level
3	1.155	1.155	1.155	1.155	1.153	1.148
4	1.499	1.496	1.492	1.481	1.463	1.425
5	1.780	1.764	1.749	1.715	1.672	1.602
6	2.011	1.973	1.944	1.887	1.822	1.729
7	2.201	2.139	2.097	2.020	1.938	1.828
8	2.358	2.274	2.221	2.126	2.032	1.909
9	2.492	2.387	2.323	2.215	2.110	1.977
10	2.606	2.482	2.410	2.290	2.176	2.036
11	2.705	2.564	2.485	2.355	2.234	2.088
12	2.791	2.636	2.550	2.412	2.285	2.134
13	2.867	2.699	2.607	2.462	2.331	2.175
14	2.935	2.755	2.659	2.507	2.371	2.213
15	2.997	2.806	2.705	2.549	2.409	2.247
16	3.052	2.852	2.747	2.585	2.443	2.279
17	3.103	2.894	2.785	2.620	2.475	2.309
18	3.149	2.932	2.821	2.651	2.504	2.335
19	3.191	2.968	2.854	2.681	2.532	2.361
20	3.230	3.001	2.884	2.709	2.557	2.385
21	3.266	3.031	2.912	2.733	2.580	2.408
22	3.300	3.060	2.939	2.758	2.603	2.429
23	3.332	3.087	2.963	2.781	2.624	2.448
24	3.362	3.112	2.987	2.802	2.644	2.467
25	3.389	3.135	3.009	2.822	2.663	2.486
26	3.415	3.157	3.029	2.841	2.681	2.502
27	3.440	3.178	3.049	2.859	2.698	2.519
28	3.464	3.199	3.068	2.876	2.714	2.534
29	3.486	3.218	3.085	2.893	2.730	2.549
30	3.507	3.236	3.103	2.908	2.745	2.563
31	3.528	3.253	3.119	2.924	2.759	2.577
32	3.546	3.270	3.135	2.938	2.773	2.591
33	3.565	3.286	3.150	2.952	2.786	2.604
34	3.582	3.301	3.164	2.965	2.799	2.616
35	3.599	3.316	3.178	2.979	2.811	2.628
36	3.616	3.330	3.191	2.991	2.823	2.639
37	3.631	3.343	3.204	3.003	2.835	2.650
38	3.646	3.356	3.216	3.014	2.846	2.661
39	3.660	3.369	3.228	3.025	2.857	2.671
40	3.673	3.381	3.240	3.036	2.866	2.682

TABLE I (CONTINUED)

Table of Critical Values for T(One-Sided Test of T_1 or T_n)
When the Standard Deviation is Calculated from the Same Samples

No. Obs. n	Upper .1% Sig. Level	Upper .5% Sig. Level	Upper 1% Sig. Level	Upper 2.5% Sig. Level	Upper 5% Sig. Level	Upper 10% Sig. Level
41	3.687	3.393	3.251	3.046	2.877	2.692
42	3.700	3.404	3.261	3.057	2.887	2.700
43	3.712	3.415	3.271	3.067	2.896	2.710
44	3.724	3.425	3.282	3.075	2.905	2.719
45	3.736	3.435	3.292	3.085	2.914	2.727
46	3.747	3.445	3.302	3.094	2.923	2.736
47	3.757	3.455	3.310	3.103	2.931	2.744
48	3.768	3.464	3.319	3.111	2.940	2.753
49	3.779	3.474	3.329	3.120	2.948	2.760
50	3.789	3.483	3.336	3.128	2.956	2.768
51	3.798	3.491	3.345	3.136	2.964	2.775
52	3.808	3.500	3.353	3.143	2.971	2.783
53	3.816	3.507	3.361	3.151	2.978	2.790
54	3.825	3.516	3.368	3.158	2.986	2.798
55	3.834	3.524	3.376	3.166	2.992	2.804
56	3.482	3.531	3.383	3.172	3.000	2.811
57	3.851	3.539	3.391	3.180	3.006	2.818
58	3.858	3.546	3.397	3.186	3.013	2.824
59	3.867	3.553	3.405	3.193	3.019	2.831
60	3.874	3.560	3.411	3.199	3.025	2.837
61	3.882	3.566	3.418	3.205	3.032	2.842
62	3.889	3.573	3.424	3.212	3.037	2.849
63	3.896	3.579	3.430	3.218	3.044	2.854
64	3.903	3.586	3.437	3.224	3.049	2.860
65	3.910	3.592	3.442	3.230	3.055	2.866
66	3.917	3.598	3.449	3.235	3.061	2.871
67	3.923	3.605	3.454	3.241	3.066	2.877
68	3.930	3.610	3.460	3.246	3.071	2.883
69	3.936	3.617	3.466	3.252	3.076	2.888
70	3.942	3.622	3.471	3.257	3.082	2.893
71	3.948	3.627	3.476	3.262	3.087	2.897
72	3.954	3.633	3.482	3.267	3.092	2.903
73	3.960	3.638	3.487	3.272	3.098	2.908
74	3.965	3.643	3.492	3.278	3.102	2.912
75	3.971	3.648	3.496	3.282	3.107	2.917
76	3.977	3.654	3.502	3.287	3.111	2.922
77	3.982	3.658	3.507	3.291	3.117	2.927
78	3.987	3.663	3.511	3.297	3.121	2.931
79	3.992	3.669	3.516	3.301	3.125	2.935
80	3.998	3.673	3.521	3.305	3.130	2.940

TABLE I (CONTINUED)

Table of Critical Values for T(One-Sided Test of T_1 or T_n) When the Standard Deviation is Calculated from the Same Samples

No. Obs. n	Upper .1% Sig. Level	Upper .5% Sig. Level	Upper 1% Sig. Level	Upper 2.5% Sig. Level	Upper 5% Sig. Level	Upper 10% Sig. Level
81	4.002	3.677	3.525	3.309	3.134	2.945
82	4.007	3.682	3.529	3.315	3.139	2.949
83	4.012	3.687	3.534	3.319	3.143	2.953
84	4.017	3.691	3.539	3.323	3.147	2.957
85	4.021	3.695	3.543	3.327	3.151	2.961
86	4.026	3.699	3.547	3.331	3.155	2.966
87	4.031	3.704	3.551	3.335	3.160	2.970
88	4.035	3.708	3.555	3.339	3.163	2.973
89	4.039	3.712	3.559	3.343	3.167	2.977
90	4.044	3.716	3.563	3.347	3.171	2.981
91	4.049	3.720	3.567	3.350	3.174	2.984
92	4.053	3.725	3.570	3.355	3.179	2.989
93	4.057	3.728	3.575	3.358	3.182	2.993
94	4.060	3.732	3.579	3.362	3.186	2.996
95	4.064	3.736	3.582	3.365	3.189	3.000
96	4.069	3.739	3.586	3.369	3.193	3.003
97	4.073	3.744	3.589	3.372	3.196	3.006
98	4.076	3.747	3.593	3.377	3.201	3.011
99	4.080	3.750	3.597	3.380	3.204	3.014
100	4.084	3.754	3.600	3.383	3.207	3.017
101	4.088	3.757	3.603	3.386	3.210	3.021
102	4.092	3.760	3.607	3.390	3.214	3.024
103	4.095	3.765	3.610	3.393	3.217	3.027
104	4.098	3.768	3.614	3.397	3.220	3.030
105	4.102	3.771	3.617	3.400	3.224	3.033
106	4.105	3.774	3.620	3.403	3.227	3.037
107	4.109	3.777	3.623	3.406	3.230	3.040
108	4.112	3.780	3.626	3.409	3.233	3.043
109	4.116	3.784	3.629	3.412	3.236	3.046
110	4.119	3.787	3.632	3.415	3.239	3.049
111	4.122	3.790	3.636	3.418	3.242	3.052
112	4.125	3.793	3.639	3.422	3.245	3.055
113	4.129	3.796	3.642	3.424	3.248	3.058
114	4.132	3.799	3.645	3.427	3.251	3.061
115	4.135	3.802	3.647	3.430	3.254	3.064
116	4.138	3.805	3.650	3.433	3.257	3.067
117	4.141	3.808	3.653	3.435	3.259	3.070

TABLE I (CONTINUED)

Table of Critical Values for T(One-Sided Test of T_1 or T_n)
When the Standard Deviation is Calculated from the Same Samples

No. Obs. n	Upper .1% Sig. Level	Upper .5% Sig. Level	Upper 1% Sig. Level	Upper 2.5% Sig. Level	Upper 5% Sig. Level	Upper 10% Sig. Level
118	4.144	3.811	3.656	3.438	3.262	3.073
119	4.146	3.814	3.659	3.441	3.265	3.075
120	4.150	3.817	3.662	3.444	3.267	3.078
121	4.153	3.819	3.665	3.447	3.270	3.081
122	4.156	3.822	3.667	3.450	3.274	3.083
123	4.159	3.824	3.670	3.452	3.276	3.086
124	4.161	3.827	3.672	3.455	3.279	3.089
125	4.164	3.831	3.675	3.457	3.281	3.092
126	4.166	3.833	3.677	3.460	3.284	3.095
127	4.169	3.836	3.680	3.462	3.286	3.097
128	4.173	3.838	3.683	3.465	3.289	3.100
129	4.175	3.840	3.686	3.467	3.291	3.102
130	4.178	3.843	3.688	3.470	3.294	3.104
131	4.180	3.845	3.690	3.473	3.296	3.107
132	4.183	3.848	3.693	3.475	3.298	3.109
133	4.185	3.850	3.695	3.478	3.302	3.112
134	4.188	3.853	3.697	3.480	3.304	3.114
135	4.190	3.856	3.700	3.482	3.306	3.116
136	4.193	3.858	3.702	3.484	3.309	3.119
137	4.196	3.860	3.704	3.487	3.311	3.122
138	4.198	3.863	3.707	3.489	3.313	3.124
139	4.200	3.865	3.710	3.491	3.315	3.126
140	4.203	3.867	3.712	3.493	3.318	3.129
141	4.205	3.869	3.714	3.497	3.320	3.131
142	4.207	3.871	3.716	3.499	3.322	3.133
143	4.209	3.874	3.719	3.501	3.324	3.135
144	4.212	3.876	3.721	3.503	3.326	3.138
145	4.214	3.879	3.723	3.505	3.328	3.140
146	4.216	3.881	3.725	3.507	3.331	142
147	4.219	3.883	3.727	3.509	3.334	3.1[illegible]4

TABLE 2 - DIXON CRITERIA FOR TESTING OF EXTREME OBSERVATION (SINGLE SAMPLE)[a]

n	Criterion	Significance Level		
		10 %	5 %	1 %
3	$r_{10} = (x_2 - x_1)/(x_n - x_1)$ if smallest	0.886	0.941	0.988
4	value is suspected;	0.679	0.765	0.889
5	$= (x_n - x_{n-1})/(x_n - x_1)$ if	0.557	0.642	0.780
6	largest value is suspected.	0.482	0.560	0.698
7		0.434	0.507	0.637
8	$r_{11} = (x_2 - x_1)/(x_{n-1} - x_1)$ if	0.479	0.554	0.683
9	smallest value is suspected;	0.441	0.512	0.635
10	$= (x_n - x_{n-1})/(x_n - x_2)$ if	0.409	0.477	0.597
	largest value is suspected.			
11	$r_{21} = (x_3 - x_1)/(x_{n-1} - x_1)$ if	0.517	0.576	0.679
12	smallest value is suspected;	0.490	0.546	0.642
13	$= (x_n - x_{n-2})/(x_n - x_2)$ if	0.467	0.521	0.615
	largest value is suspected.			
14	$r_{22} = (x_3 - x_1)/(x_{n-2} - x_1)$ if	0.492	0.546	0.641
15	smallest value is suspected;	0.472	0.525	0.616
16	$= (x_n - x_{n-2})/(x_n - x_3)$ if	0.454	0.507	0.595
17	largest value is suspected.	0.438	0.490	0.577
18		0.424	0.475	0.561
19		0.412	0.462	0.547
20		0.401	0.450	0.535
21		0.391	0.440	0.524
22		0.382	0.430	0.514
23		0.374	0.421	0.505
24		0.367	0.413	0.497
25		0.360	0.406	0.489

[a] $x_1 \leq x_2 \leq \ldots \leq x_n$. (See Ref(7), Appendix.)

TABLE 3

CRITICAL VALUES FOR w/s (RATIO OF RANGE TO SAMPLE STANDARD DEVIATION)[a]

Number of Observations n	5 Percent Significance Level	1 Percent Significance Level	0.5 Percent Significance Level
3	2.00	2.00	2.00
4	2.43	2.44	2.45
5	2.75	2.80	2.81
6	3.01	3.10	3.12
7	3.22	3.34	3.37
8	3.40	3.54	3.58
9	3.55	3.72	3.77
10	3.68	3.88	3.94
11	3.80	4.01	4.08
12	3.91	4.13	4.21
13	4.00	4.24	4.32
14	4.09	4.34	4.43
15	4.17	4.43	4.53
16	4.24	4.51	4.62
17	4.31	4.59	4.69
18	4.38	4.66	4.77
19	4.43	4.73	4.84
20	4.49	4.79	4.91
30	4.89	5.25	5.39
40	5.15	5.54	5.69
50	5.35	5.77	5.91
60	5.50	5.93	6.09
80	5.73	6.18	6.35
100	5.90	6.36	6.54
150	6.18	6.64	6.84
200	6.38	6.85	7.03
500	6.94	7.42	7.60
1000	7.33	7.80	7.99

[a]See Ref (5), where:

$w = x_n - x_1 \quad , \quad x_1 \leq x_2 \leq \ldots \leq x_n$

$s = \sqrt{\Sigma(x_i - \bar{x})^2/(n-1)}$

TABLE 4 *

TABLE OF CRITICAL VALUES FOR $S^2_{n-1,n}/S^2$ OR $S^2_{1,2}/S^2$ FOR SIMULTANEOUSLY TESTING THE TWO LARGEST OR TWO SMALLEST OBSERVATIONS

No. of Obs. n	Lower .1% Sig. Level	Lower .5% Sig. Level	Lower 1% Sig. Level	Lower 2.5% Sig. Level	Lower 5% Sig. Level	Lower 10% Sig. Level
4	.0000	.0000	.0000	.0002	.0008	.0031
5	.0003	.0018	.0035	.0090	.0183	.0376
6	.0039	.0116	.0186	.0349	.0564	.0920
7	.0135	.0308	.0440	.0708	.1020	.1479
8	.0290	.0563	.0750	.1101	.1478	.1994
9	.0489	.0851	.1082	.1492	.1909	.2454
10	.0714	.1150	.1414	.1864	.2305	.2863
11	.0953	.1448	.1736	.2213	.2667	.3227
12	.1198	.1738	.2043	.2537	.2996	.3552
13	.1441	.2016	.2333	.2836	.3295	.3843
14	.1680	.2280	.2605	.3112	.3568	.4106
15	.1912	.2530	.2859	.3367	.3818	.4345
16	.2136	.2767	.3098	.3603	.4048	.4562
17	.2350	.2990	.3321	.3822	.4259	.4761
18	.2556	.3200	.3530	.4025	.4455	.4944
19	.2752	.3398	.3725	.4214	.4636	.5113
20	.2939	.3585	.3909	.4391	.4804	.5270
21	.3118	.3761	.4082	.4556	.4961	.5415
22	.3288	.3927	.4245	.4711	.5107	.5550
23	.3450	.4085	.4398	.4857	.5244	.5677
24	.3605	.4234	.4543	.4994	.5373	.5795
25	.3752	.4376	.4680	.5123	.5495	.5906

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \; ; \; \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \; ; \; x_1 \leq x_2 \leq \dots \leq x_n$$

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

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

*These significance levels are taken from Table II, Ref. (12). An observed ratio less than the appropriate critical ratio in this table calls for rejection of the null hypothesis.

TABLE 4 (CONTINUED)

TABLE OF CRITICAL VALUES FOR $S^2_{n-1,n}/S^2$ OR $S^2_{1,2}/S^2$ FOR SIMULTANEOUSLY TESTING THE TWO LARGEST OR TWO SMALLEST OBSERVATIONS

No. of Obs. n	Lower .1% Sig. Level	Lower .5% Sig. Level	Lower 1% Sig. Level	Lower 2.5% Sig. Level	Lower 5% Sig. Level	Lower 10% Sig. Level
26	.3893	.4510	.4810	.5245	.5609	.6011
27	.4027	.4638	.4933	.5360	.5717	.6110
28	.4156	.4759	.5050	.5470	.5819	.6203
29	.4279	.4875	.5162	.5574	.5916	.6292
30	.4397	.4985	.5268	.5672	.6008	.6375
31	.4510	.5091	.5369	.5766	.6095	.6455
32	.4618	.5192	.5465	.5856	.6178	.6530
33	.4722	.5288	.5557	.5941	.6257	.6602
34	.4821	.5381	.5646	.6023	.6333	.6671
35	.4917	.5469	.5730	.6101	.6405	.6737
36	.5009	.5554	.5811	.6175	.6474	.6800
37	.5098	.5636	.5889	.6247	.6541	.6860
38	.5184	.5714	.5963	.6316	.6604	.6917
39	.5266	.5789	.6035	.6382	.6665	.6972
40	.5345	.5862	.6104	.6445	.6724	.7025
41	.5422	.5932	.6170	.6506	.6780	.7076
42	.5496	.5999	.6234	.6565	.6834	.7125
43	.5568	.6064	.6296	.6621	.6886	.7172
44	.5637	.6127	.6355	.6676	.6936	.7218
45	.5704	.6188	.6412	.6728	.6985	.7261
46	.5768	.6246	.6468	.6779	.7032	.7304
47	.5831	.6303	.6521	.6828	.7077	.7345
48	.5892	.6358	.6573	.6876	.7120	.7384
49	.5951	.6411	.6623	.6921	.7163	.7422
50	.6008	.6462	.6672	.6966	.7203	.7459
51	.6063	.6512	.6719	.7009	.7243	.7495
52	.6117	.6560	.6765	.7051	.7281	.7529
53	.6169	.6607	.6809	.7091	.7319	.7563
54	.6220	.6653	.6852	.7130	.7355	.7595
55	.6269	.6697	.6894	.7168	.7390	.7627
56	.6317	.6740	.6934	.7205	.7424	.7658
57	.6364	.6782	.6974	.7241	.7456	.7687
58	.6410	.6823	.7012	.7276	.7489	.7716
59	.6454	.6862	.7049	.7310	.7520	.7744
60	.6497	.6901	.7086	.7343	.7550	.7772
61	.6539	.6938	.7121	.7375	.7580	.7798
62	.6580	.6975	.7155	.7406	.7608	.7824
63	.6620	.7010	.7189	.7437	.7636	.7850
64	.6658	.7045	.7221	.7467	.7664	.7874
65	.6696	.7079	.7253	.7496	.7690	.7898
66	.6733	.7112	.7284	.7524	.7716	.7921
67	.6770	.7144	.7314	.7551	.7741	.7944

TABLE 4 (CONTINUED)

TABLE OF CRITICAL VALUES FOR $S^2_{n-1,n}/S^2$ OR $S^2_{1,2}/S^2$ FOR SIMULTANEOUSLY TESTING THE TWO LARGEST OR TWO SMALLEST OBSERVATIONS

No. of Obs. n	Lower .1% Sig. Level	Lower .5% Sig. Level	Lower 1% Sig. Level	Lower 2.5% Sig. Level	Lower 5% Sig. Level	Lower 10% Sig. Level
68	.6805	.7175	.7344	.7578	.7766	.7966
69	.6839	.7206	.7373	.7604	.7790	.7988
70	.6873	.7236	.7401	.7630	.7813	.8009
71	.6906	.7265	.7429	.7655	.7836	.8030
72	.6938	.7294	.7455	.7679	.7859	.8050
73	.6970	.7322	.7482	.7703	.7881	.8070
74	.7000	.7349	.7507	.7727	.7902	.8089
75	.7031	.7376	.7532	.7749	.7923	.8108
76	.7060	.7402	.7557	.7772	.7944	.8127
77	.7089	.7427	.7581	.7794	.7964	.8145
78	.7117	.7453	.7605	.7815	.7983	.8162
79	.7145	.7477	.7628	.7836	.8002	.8180
80	.7172	.7501	.7650	.7856	.8021	.8197
81	.7199	.7525	.7672	.7876	.8040	.8213
82	.7225	.7548	.7694	.7896	.8058	.8230
83	.7250	.7570	.7715	.7915	.8075	.8245
84	.7275	.7592	.7736	.7934	.8093	.8261
85	.7300	.7614	.7756	.7953	.8109	.8276
86	.7324	.7635	.7776	.7971	.8126	.8291
87	.7348	.7656	.7796	.7989	.8142	.8306
88	.7371	.7677	.7815	.8006	.8158	.8321
89	.7394	.7697	.7834	.8023	.8174	.8335
90	.7416	.7717	.7853	.8040	.8190	.8349
91	.7438	.7736	.7871	.8057	.8205	.8362
92	.7459	.7755	.7889	.8073	.8220	.8376
93	.7481	.7774	.7906	.8089	.8234	.8389
94	.7501	.7792	.7923	.8104	.8248	.8402
95	.7522	.7810	.7940	.8120	.8263	.8414
96	.7542	.7828	.7957	.8135	.8276	.8427
97	.7562	.7845	.7973	.8149	.8290	.8439
98	.7581	.7862	.7989	.8164	.8303	.8451
99	.7600	.7879	.8005	.8178	.8316	.8463
100	.7619	.7896	.8020	.8192	.8329	.8475
101	.7637	.7912	.8036	.8206	.8342	.8486
102	.7655	.7928	.8051	.8220	.8354	.8497
103	.7673	.7944	.8065	.8233	.8367	.8508
104	.7691	.7959	.8080	.8246	.8379	.8519
105	.7708	.7974	.8094	.8259	.8391	.8530
106	.7725	.7989	.8108	.8272	.8402	.8541

TABLE 4 (CONTINUED)

TABLE OF CRITICAL VALUES FOR $S^2_{n-1,n}/S^2$ OR $S^2_{1,2}/S^2$ FOR SIMULTANEOUSLY TESTING THE TWO LARGEST OR TWO SMALLEST OBSERVATIONS

No. of Obs. n	Lower .1% Sig. Level	Lower .5% Sig. Level	Lower 1% Sig. Level	Lower 2.5% Sig. Level	Lower 5% Sig. Level	Lower 10% Sig. Level
107	.7742	.8004	.8122	.8284	.8414	.8551
108	.7758	.8018	.8136	.8297	.8425	.8563
109	.7774	.8033	.8149	.8309	.8436	.8571
110	.7790	.8047	.8162	.8321	.8447	.8581
111	.7806	.8061	.8175	.8333	.8458	.8591
112	.7821	.8074	.8188	.8344	.8469	.8600
113	.7837	.8088	.8200	.8356	.8479	.8610
114	.7852	.8101	.8213	.8367	.8489	.8619
115	.7866	.8114	.8225	.8378	.8500	.8628
116	.7881	.8127	.8237	.8389	.8510	.8637
117	.7895	.8139	.8249	.8400	.8519	.8646
118	.7909	.8152	.8261	.8410	.8529	.8655
119	.7923	.8164	.8272	.8421	.8539	.8664
120	.7937	.8176	.8284	.8431	.8548	.8672
121	.7951	.8188	.8295	.8441	.8557	.8681
122	.7964	.8200	.8306	.8451	.8567	.8689
123	.7977	.8211	.8317	.8461	.8576	.8697
124	.7990	.8223	.8327	.8471	.8585	.8705
125	.8003	.8234	.8338	.8480	.8593	.8713
126	.8016	.8245	.8348	.8490	.8602	.8721
127	.8028	.8256	.8359	.8499	.8611	.8729
128	.8041	.8267	.8369	.8508	.8619	.8737
129	.8053	.8278	.8379	.8517	.8627	.8744
130	.8065	.8288	.8389	.8526	.8636	.8752
131	.8077	.8299	.8398	.8535	.8644	.8759
132	.8088	.8309	.8408	.8544	.8652	.8766
133	.8100	.8319	.8418	.8553	.8660	.8773
134	.8111	.8329	.8427	.8561	.8668	.8780
135	.8122	.8339	.8436	.8570	.8675	.8787
136	.8134	.8349	.8445	.8578	.8683	.8794
137	.8145	.8358	.8454	.8586	.8690	.8801
138	.8155	.8368	.8463	.8594	.8698	.8808
139	.8166	.8377	.8472	.8602	.8705	.8814
140	.8176	.8387	.8481	.8610	.8712	.8821
141	.8187	.8396	.8489	.8618	.8720	.8827
142	.8197	.8405	.8498	.8625	.8727	.8834
143	.8207	.8414	.8506	.8633	.8734	.8840
144	.8218	.8423	.8515	.8641	.8741	.8846
145	.8227	.8431	.8523	.8648	.8747	.8853
146	.8237	.8440	.8531	.8655	.8754	.8859
147	.8247	.8449	.8539	.8663	.8761	.8865
148	.8256	.8457	.8547	.8670	.8767	.8871
149	.8266	.8465	.8555	.8677	.8774	.8877

TABLE 5

CRITICAL VALUES FOR T' WHEN STANDARD DEVIATION s_ν IS INDEPENDENT OF PRESENT SAMPLE[a]

$$T' = \frac{x_n - \bar{x}}{s_\nu}, \text{ or } \frac{\bar{x} - x_1}{s_\nu}$$

ν=d.f.	n = 3	4	5	6	7	8	9	10	12
				1 percentage point					
10	2.78	3.10	3.32	3.48	3.62	3.73	3.82	3.90	4.04
11	2.72	3.02	3.24	3.39	3.52	3.63	3.72	3.79	3.93
12	2.67	2.96	3.17	3.32	3.45	3.55	3.64	3.71	3.84
13	2.63	2.92	3.12	3.27	3.38	3.48	3.57	3.64	3.76
14	2.60	2.88	3.07	3.22	3.33	3.43	3.51	3.58	3.70
15	2.57	2.84	3.03	3.17	3.29	3.38	3.46	3.53	3.65
16	2.54	2.81	3.00	3.14	3.25	3.34	3.42	3.49	3.60
17	2.52	2.79	2.97	3.11	3.22	3.31	3.38	3.45	3.56
18	2.50	2.77	2.95	3.08	3.19	3.28	3.35	3.42	3.53
19	2.49	2.75	2.93	3.06	3.16	3.25	3.33	3.39	3.50
20	2.47	2.73	2.91	3.04	3.14	3.23	3.30	3.37	3.47
24	2.42	2.68	2.84	2.97	3.07	3.16	3.23	3.29	3.38
30	2.38	2.62	2.79	2.91	3.01	3.08	3.15	3.21	3.30
40	2.34	2.57	2.73	2.85	2.94	3.02	3.08	3.13	3.22
60	2.29	2.52	2.68	2.79	2.88	2.95	3.01	3.06	3.15
120	2.25	2.48	2.62	2.73	2.82	2.89	2.95	3.00	3.08
∞	2.22	2.43	2.57	2.68	2.76	2.83	2.88	2.93	3.01
				5 percentage points					
10	2.01	2.27	2.46	2.60	2.72	2.81	2.89	2.96	3.08
11	1.98	2.24	2.42	2.56	2.67	2.76	2.84	2.91	3.03
12	1.96	2.21	2.39	2.52	2.63	2.72	2.80	2.87	2.98
13	1.94	2.19	2.36	2.50	2.60	2.69	2.76	2.83	2.94
14	1.93	2.17	2.34	2.47	2.57	2.66	2.74	2.80	2.91
15	1.91	2.15	2.32	2.45	2.55	2.64	2.71	2.77	2.88
16	1.90	2.14	2.31	2.43	2.53	2.62	2.69	2.75	2.86
17	1.89	2.13	2.29	2.42	2.52	2.60	2.67	2.73	2.84
18	1.88	2.11	2.28	2.40	2.50	2.58	2.65	2.71	2.82
19	1.87	2.11	2.27	2.39	2.49	2.57	2.64	2.70	2.80
20	1.87	2.10	2.26	2.38	2.47	2.56	2.63	2.68	2.78
24	1.84	2.07	2.23	2.34	2.44	2.52	2.58	2.64	2.74
30	1.82	2.04	2.20	2.31	2.40	2.48	2.54	2.60	2.69
40	1.80	2.02	2.17	2.28	2.37	2.44	2.50	2.56	2.65
60	1.78	1.99	2.14	2.25	2.33	2.41	2.47	2.52	2.61
120	1.76	1.96	2.11	2.22	2.30	2.37	2.43	2.48	2.57
∞	1.74	1.94	2.08	2.18	2.27	2.33	2.39	2.44	2.52

[a]The percentage points are reproduced from Ref (4).

TABLE 6

STANDARDIZATION OF SODIUM HYDROXIDE SOLUTIONS AS DETERMINED BY PLANT LABORATORIES

Standard Used: Potassium Acid Phthalate (P.A.P.)

Laboratory	(P.A.P. - 0.096000) x 10^3	Sums	Averages	Deviation of Average from Grand Average
1	1.893 1.972 1.876	5.741	1.914	+0.043
2	2.046 1.851 1.949	5.846	1.949	+0.078
3	1.874 1.792 1.829	5.495	1.832	-0.039
4	1.861 1.998 1.983	5.842	1.947	+0.076
5	1.922 1.881 1.850	5.653	1.884	+0.013
6	2.082 1.958 2.029	6.069	2.023	+0.152
7	1.992 1.980 2.066	6.038	2.013	+0.142
8	2.050 2.181 1.903	6.134	2.045	+0.174
9	1.831 1.883 1.855	5.569	1.856	-0.015
10	0.735 0.722 0.777	2.234	0.745	-1.126
11	2.064 1.794 1.891	5.749	1.916	+0.045
12	2.475 2.403 2.102	6.980	2.327	+0.456
Grand Sum		67.350		
Grand Average			1.871	

TABLE 7

ANALYSIS OF VARIANCE

Source of Variation	Degrees of Freedom (d.f.)	Sum of Squares (SS)	Mean Square (MS)	F-ratio
Between laboratories Within laboratories	11 24	4.70180 0.21103	0.4274 0.008793	F = 48.61 (highly significant)
Total	35	4.91283		

TABLE 8

ANALYSIS OF VARIANCE
(Omitting Labs 10 and 12)

Source of Variation	d.f.	SS	MS	F-ratio
Between laboratories Within laboratories	9 20	0.13889 0.13107	0.01543 0.00655	F = 2.36 $F_{0.05}(9,20) = 2.40$ $F_{0.01}(9,20) = 3.45$
Total	29	0.26996		

TABLE 9

CRITICAL VALUES OF $T'_{1\infty}$ AND $T'_{n\infty}$ WHEN THE POPULATION STANDARD DEVIATION σ IS KNOWN[a]

Number of Observations n	5 Percent Significance Level	1 Percent Significance Level	0.5 Percent Significance Level
2	1.39	1.82	1.99
3	1.74	2.22	2.40
4	1.94	2.43	2 62
5	2.08	2.57	2.76
6	2.18	2.68	2.87
7	2.27	2.76	2.95
8	2.33	2.83	3.02
9	2.39	2.88	3.07
10	2.44	2.93	3.12
11	2.48	2.97	3.16
12	2.52	3.01	3.20
13	2.56	3.04	3.23
14	2.59	3.07	3.26
15	2.62	3.10	3.29
16	2.64	3.12	3.31
17	2.67	3.15	3.33
18	2.69	3.17	3.36
19	2.71	3.19	3.38
20	2.73	3.21	3.39
21	2.75	3.22	3.41
22	2.77	3.24	3.42
23	2.78	3.26	3.44
24	2.80	3.27	3.45
25	2.81	3.28	3.46

$$x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n$$

$$T'_{1\infty} = (\bar{x} - x_1)/\sigma \; ; \; T'_{n\infty} = (x_n - \bar{x})/\sigma$$

[a]This table is taken from Ref (10).

TABLE 10

MEASUREMENTS, μm

x-Coordinate			y-Coordinate		
Position 1	Position 1 + 180 deg	Δx	Position 1	Position 1 + 180 deg	Δy
-53011	-53004	-7	70263	70258	+5
-38112	-38103	-9	-39729	-39723	-6
- 2804	- 2828	+24	81162	81140	+22
18473	18467	+6	41477	41485	-8
25507	25497	+10	1082	1076	+6
87736	87739	-3	-7442	-7434	-8

TABLE 11

Critical Values for $L_{k,\ \alpha} = 0.01$

n/k	1	1*	2	2**	3	4	5	6	7	8	9	10
3	.000	.000										
4	.011	.010	.000	.000								
5	.045	.044	.004	.004								
6	.091	.093	.021	.019	.002							
7	.148	.145	.047	.044	.010							
8	.202	.195	.076	.075	.028	.008						
9	.235	.241	.112	.108	.048	.018						
10	.280	.283	.142	.142	.070	.032	.012					
11	.327	.321	.178	.174	.098	.052	.026					
12	.371	.355	.208	.204	.120	.070	.038	.019				
13	.400	.386	.233	.233	.147	.094	.056	.033				
14	.424	.414	.267	.261	.172	.113	.072	.406	.027			
15	.450	.440	.294	.286	.194	.132	.090	.057	.037			
16	.473	.463	.311	.310	.219	.151	.108	.072	.049	.030		
17	.480	.485	.338	.332	.237	.171	.126	.091	.064	.044		
18	.502	.504	.358	.353	.260	192	.140	.104	.076	.053	.036	
19	.508	.522	.366	.373	.272	.201	.154	.118	.088	.064	.046	
20	.533	.539	.387	.391	.300	.231	.175	.136	.104	.078	.058	.042
25	.603		.468		.377	.308	.246	.204	.168	.144	.112	.092
30	.650		.526		.434	.369	.312	.268	.229	.196	.166	.142
35	.690		.574		.484	.418	.364	.321	.282	.250	.220	.194
40	.722		.608		.522	.460	.408	.364	.324	.292	.262	.234
45	.745		.636		.558	.498	.444	.399	.361	.328	.296	.270
50	.768		.668		.592	.531	.483	.438	.400	.368	.336	.308

* From Grubbs (1950, Table I).

** From Grubbs (1950, Table V).

(This table is taken from Tietjen and Moore, Reference 19)

TABLE 11

Critical Values for $L_{k,\alpha} = 0.025$

n/k	1	1*	2	2**	3	4	5	6	7	8	9	10
3	.001	.001	.000	.000								
4	.025	.025	.000	.000								
5	.084	.081	.011	.009								
6	.146	.145	.034	.035	.005							
7	.209	.207	.076	.071	.021							
8	.262	.262	.115	.110	.045	.013						
9	.308	.310	.150	.149	.073	.030						
10	.350	.353	.188	.187	.100	.052	.023					
11	.366	.390	.225	.221	.129	.074	.040					
12	.440	.423	.268	.254	.162	.096	.057	.031				
13	.462	.453	.292	.284	.184	.122	.077	.047				
14	.493	.479	.317	.311	.214	.145	.098	.063	.038			
15	.498	.503	.341	.337	.239	.167	.111	.078	.051			
16	.537	.525	.372	.360	.261	.185	.137	.096	.065	.045		
17	.552	.544	.388	.382	.282	.208	.156	.117	.082	.058		
18	.570	.562	.406	.403	.299	.226	.171	.129	.095	.068	.048	
19	.573	.579	.416	.421	.311	.243	.189	.145	.108	.080	.059	
20	.595	.594	.442	.439	.341	.265	.209	.165	.128	.098	.073	.054
25	.656		.654		.516	.417	.342	.282	.233	.192	.159	.132
30	.699		.568		.479	.408	.352	.302	.261	.226	.193	.165
35	.732		.612		.527	.435	.398	.348	.308	.274	.242	.213
40	.755		.641		.561	.491	.433	.387	.348	.314	.283	.257
45	.773		.667		.592	.529	.473	.430	.391	.356	.325	.295
50	.796		.698		.622	.559	.510	.466	.428	.392	.363	.334

*From Grubbs (1950, Table I).
**From Grubbs (1950, Table V).

TABLE 11

Critical Values for $L_{k,\alpha} = 0.05$

n/k	1	1*	2	2**	3	4	5	6	7	8	9	10
3	.003	.003										
4	.051	.049	.001	.001								
5	.125	.127	.018	.018								
6	.203	.203	.055	.057	.010							
7	.273	.270	.106	.102	.032							
8	.326	.326	.146	.148	.064	.022						
9	.372	.374	.194	.191	.099	.045						
10	.418	.415	.233	.230	.129	.070	.034					
11	.454	.451	.270	.267	.162	.098	.054					
12	.489	.482	.305	.300	.196	.125	.076	.042				
13	.517	.510	.337	.330	.224	.150	.098	.060				
14	.540	.534	.363	.357	.250	.174	.122	.079	.050			
15	.556	.556	.387	.382	.276	.197	.140	.097	.066			
16	.575	.576	.410	.405	.300	.219	.159	.115	.082	.055		
17	.594	.593	.427	.426	.322	.240	.181	.136	.100	.072		
18	.608	.610	.447	.446	.337	.259	.200	.154	.116	.086	.062	
19	.624	.624	.462	.464	.354	.277	.209	.168	.130	.099	.074	
20	.639	.638	.484	.480	.377	.299	.238	.188	.150	.115	.088	.066
25	.696	.692	.550		.450	.374	.312	.262	.222	.184	.154	.126
30	.730		.599		.506	.434	.376	.327	.283	.245	.212	.183
35	.762		.642		.554	.482	.424	.376	.334	.297	.264	.235
40	.784		.672		.588	.523	.468	.421	.378	.342	.310	.280
45	.802		.696		.618	.556	.502	.456	.417	.382	.350	.320
50	.820		.722		.646	.588	.535	.490	.450	.414	.383	.356

*From Grubbs (1950, Table I).
**From Grubbs (1950, Table V).

TABLE 11

Critical Values for $L_{k,\alpha} = 0.10$

n/k	1	1*	2	2**	3	4	5	6	7	8	9	10
3	.011	.011										
4	.098	.098	.003	.003								
5	.200	.199	.038	.038								
6	.280	.283	.091	.092	.020							
7	.348	.350	.148	.148	.056							
8	.404	.405	.200	.199	.095	.038						
9	.448	.450	.248	.245	.134	.068						
10	.490	.488	.287	.286	.170	.098	.051					
11	.526	.520	.326	.323	.208	.128	.074					
12	.555	.548	.361	.355	.240	.159	.103	.062				
13	.578	.573	.388	.384	.270	.186	.126	.082				
14	.600	.594	.416	.411	.298	.212	.150	.104	.068			
15	.611	.613	.436	.435	.322	.236	.172	.124	.086			
16	.631	.631	.458	.456	.342	.260	.194	.144	.104	.073		
17	.648	.646	.478	.476	.364	.282	.216	.165	.125	.092		
18	.661	.660	.496	.494	.384	.302	.236	.184	.142	.108	.080	
19	.676	.673	.510	.511	.398	.316	.251	.199	.158	.124	.094	
20	.688	.685	.530	.527	.420	.339	.273	.220	.176	.140	.110	.085
25	.732	.732	.588		.489	.412	.350	.296	.251	.213	.180	.152
30	.766		.637		.523	.472	.411	.359	.316	.276	.240	.210
35	.792		.673		.586	.516	.458	.410	.365	.328	.294	.262
40	.812		.702		.622	.554	.499	.451	.408	.372	.338	.307
45	.826		.724		.648	.586	.533	.488	.447	.410	.378	.348
50	.840		.744		.673	.614	.562	.518	.477	.442	.410	.380

*From Grubbs (1950, Table I)
**From Grubbs (1950, Table V).

TABLE 12

Critical Values for $E_{k,\alpha} = 0.01$

n/k	1	2	3	4	5	6	7	8	9	10
3	.000									
4	.004	.000								
5	.029	002								
6	.068	.012	.001							
7	.110	.028	.006							
8	.156	.050	.014	.004						
9	.197	.078	.026	.009						
10	.235	.101	.018	.006						
11	.274	.134	.064	.030	.012					
12	.311	.159	.083	.042	.020	.008				
13	.337	.181	.103	.056	.031	.014				
14	.374	.207	.123	.072	.042	.022	.012			
15	.404	.238	.146	.090	.054	.032	.018			
16	.422	.263	.166	.107	.068	.040	.024	.014		
17	.440	.290	.188	.122	.079	.052	.032	.018		
18	.459	.306	.206	.141	.094	.062	.041	.026	.014	
19	.484	.323	.219	.156	.108	.074	.050	.032	.020	
20	.499	.339	.236	.170	.121	.086	.058	.040	.026	.017
25	.571	.418	.320	.245	.188	.146	.110	.087	.066	.050
30	.624	.482	.386	.308	.250	.204	.166	.132	.108	.087
35	.669	.533	.435	.364	.299	.252	.211	.177	.149	.124
40	.704	.574	.480	.408	.347	.298	.258	.220	.190	.164
45	.728	.607	.518	.446	.386	.336	.294	.258	.228	.200
50	.748	.636	.550	.482	.424	.376	.334	.297	.264	.235

(This table is taken from Tietjen and Moore, Reference 19)

TABLE 12

Critical Values for $E_{k,\alpha} = 0.05$

n/k	1	1*	2	3	4	5	6	7	8	9	10
3	.001	.001									
4	.025	.025	.001								
5	.081	.081	.010								
6	.146	.145	.034	.004							
7	.208	.207	.065	.016							
8	.265	.262	.099	.034	.010						
9	.314	.310	.137	.057	.021						
10	.356	.352	.172	.083	.037	.014					
11	.386	.390	.204	.107	.055	.026					
12	.424	.423	.234	.133	.073	.039	.018				
13	.455	.453	.262	.156	.092	.053	.028				
14	.484	.479	.293	.179	.112	.068	.039	.021			
15	.509	.503	.317	.206	.134	.084	.052	.030			
16	.526	.525	.340	.227	.153	.102	.067	.041	.024		
17	.544	.544	.362	.248	.170	.116	.078	.050	.032		
18	.562	.562	.382	.267	.187	.132	.091	.062	.041	.026	
19	.581	.579	.398	.287	.203	.146	.105	.074	.050	.033	
20	.597	.594	.416	.302	.221	.163	.119	.085	.059	.041	.028
25	.652	.654	.493	.381	.298	.236	.186	.146	.114	.089	.068
30	.698		.549	.443	.364	.298	.246	.203	.166	.137	.112
35	.732		.596	.495	.417	.351	.298	.254	.214	.181	.154
40	.758		.629	.534	.458	.395	.343	.297	.259	.223	.195
45	.778		.658	.567	.492	.433	.381	.337	.299	.263	.233
50	.797		.684	.599	.529	.468	.417	.373	.334	.299	.268

*From Grubbs, Ref. (11), 1950.

TABLE 12

Critical Values for $E_{k,\alpha} - 0.10$

n/k	1	1*	2	3	4	5	6	7	8	9	10
3	.003	.003									
4	.050	.049	.002								
5	.127	.127	.022								
6	.204	.203	.056	.009							
7	.268	.270	.094	.027							
8	.328	.326	.137	.053	.016						
9	.377	.374	.175	.080	.032						
10	.420	.415	.214	.108	.052	.022					
11	.449	.451	.250	.138	.073	.036					
12	.485	.482	.278	.162	.094	.052	.026				
13	.510	.510	.309	.189	.116	.068	.038				
14	.538	.534	.337	.216	.138	.086	.052	.029			
15	.558	.556	.360	.240	.160	.105	.067	.040			
16	.578	.576	.384	.263	.182	.122	.082	.053	.032		
17	.594	.593	.406	.284	.198	.140	.095	.064	.042		
18	.610	.610	.424	.304	.217	.156	.110	.076	.051	.034	
19	.629	.624	.442	.322	.234	.172	.124	.089	.062	.042	
20	.644	.638	.460	.338	.252	188	.138	.102	.072	.051	.035
25	.693	.692	.528	.417	.331	.264	.210	.168	.132	.103	.080
30	.730		.582	.475	.391	.325	.270	.224	.186	.154	.126
35	.763		.624	.523	.443	.379	.324	.276	.236	.202	.172
40	.784		.657	.562	.486	.422	.367	.320	.278	.243	.212
45	.803		.684	.593	.522	.459	.406	.360	.320	.284	.252
50	.820		.708	.622	.552	.492	.440	.396	.355	.319	.287

*From Grubbs, Ref. (11), 1950.

REFERENCES

(1) Anscombe, F. J., "Rejection of Outliers", *Technometrics*, Vol 2, No. 2, 1960, pp 123-147.

(2) Chauvenet, W., *A Manual of Spherical and Practical Astronomy*, Vol 2, Fifth Edition, 1876.

(3) Chew, Victor, "Tests for the Rejection of Outlying Observations", RCA Systems Analysis Technical Memorandum No. 64-7, 31 Dec 1964, Patrick Air Force Base, Fla.

(4) David, H. A., "Revised Upper Percentage Points of the Extreme Studentized Deviate from the Sample Mean", *Biometrika*, Vol 43, 1956, pp 449-451.

(5) David, H. A., Hartley, H. O., and Pearson, E. S., "The Distribution of the Ratio in a Single Sample of Range to Standard Deviation", *Biometrika*, Vol 41, 1954, pp 482-493.

(6) David, H. A., and Quesenberry, C. P., "Some Tests for Outliers", Technical Report No. 47, OOR(ARO) Project No. 1166, Virginia Polytechnic Inst., Blacksburg, Va.

(7) Dixon, W. J., "Processing Data for Outliers", *Biometrics*, March 1953, Vol 9, No. 1, pp 74-89.

(8) Ferguson, T. S., "On the Rejection of Outliers" *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, edited by Jerzy Neyman, University of California Press, Berkeley and Los Angeles, Calif., 1961.

(9) Ferguson, T. S., "Rules for Rejection of Outliers", *Revue Inst. Int. de Stat.*, Vol 3, 1961, pp 29-43.

(10) Grubbs, F. E., "Procedures for Detecting Outlying Observations in Samples", *Technometrics*, Vol 11, No. 1, February 1969, pp 1-21.

(11) Grubbs, F. E., "Sample Criteria for Testing Outlying Observations", *Annals of Mathematical Statistics*, Vol 21, March 1950, pp 27-58.

REFERENCES

(12) Grubbs, F. E. and Beck, Glenn, "Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations", *Technometrics*, Vol 14, No. 4, November 1972, pp 847-854.

(13) Greenhouse, S. W., Halperin M., and Cornfield, J., "Tables of Percentage Points for the Studentized Maximum Absolute Deviation in Normal Samples", *Journal of the American Statistical Association*, Vol 50, No. 269, 1955, pp 185-195.

(14) Kruskal, W. H., "Some Remarks on Wild Observations", *Technometrics*, Vol 2, No. 1, 1960, pp 1-3.

(15) Kudo, A., "On the Testing of Outlying Observations", *Sankhya, The Indian Journal of Statistics*, Vol 17, Part 1, June 1956, pp 67-76.

(16) Proschan, F., "Testing Suspected Observations", *Industrial Quality Control*, Vol XIII, No. 7, January 1957, pp 14-19.

(17) Sarhan, A. E., and Greenberg, B. G., Editors, *Contributions to Order Statistics* , John Wiley and Sons, Inc., New York, 1962.

(18) Thompson, W. R., "On a Criterion for the Rejection of Observations and the Distribution of the Ratio of the Deviation to the Sample Standard Deviation" *The Annals of Mathematical Statistics*, Vol 6, 1935, pp 214-219.

(19) Tietjen, G. L. and Moore, R. H., "Some Grubbs-Type Statistics for the Detection of Several Outliers", *Technometrics*, Vol 14, No. 3, August 1972, pp 583-597.