NOTICE: This document is a TAPPI Standard in ballot. Although available for public viewing, it is still under TAPPI's copyright and may not be reproduced or distributed without permission of TAPPI. This document is current under review to be maintained as a TAPPI Standard.

WI 240800.01

T 1205

BALLOT NO. 2 - SARG

DRAFT NO. 01

DATE October 22, 2024

WORKING GROUP
CHAIR N/A

SUBJECT
CATEGORY
Process and Product Quality (General)

RELATED
METHODS See "Additional Information"

CAUTION:
This Test Method may include safety precautions which are believed to be appropriate at the time of publication of the method. The intent of these is to alert the user of the method to safety issues related to such use. The user is responsible for determining that the safety precautions are complete and are appropriate to their use of the method, and for ensuring that suitable safety practices have not changed since publication of the method. This method may require the use, disposal, or both, of chemicals which may present serious health hazards to humans. Procedures for the handling of such substances are set forth on Safety Data Sheets which must be developed by all manufacturers and importers of potentially hazardous chemicals and maintained by all distributors of potentially hazardous chemicals. Prior to the use of this method, the user must determine whether any of the chemicals to be used or disposed of are potentially hazardous and, if so, must follow strictly the procedures specified by both the manufacturer, as well as local, state, and federal authorities for safe use and disposal of these chemicals.

Dealing with Suspect (Outlying) Test Determinations

(Five-year review of Standard Practice T 1205 sp-20)

1. Scope

1.1 This TAPPI Standard Practice provides a procedure for judging whether suspect test determinations should be investigated further for possible rejection. A suspect determination (apparent outlier) is one that appears to deviate markedly from other determinations on the same sample of material. An outlying determination (outlier) is a suspect determination for which the deviation has, in fact, been found to be significant using an appropriate statistical test.

1.2 Formal treatment of suspect test determinations, as specified in this document, is necessary only in critical situations (e.g., very critical research) or when required by a product specification or an official test method.

1.2.1 Formal treatment of suspect test determinations and test results is highly desirable in studies establishing the repeatability and reproducibility of a test method (see TAPPI T 1200 "Interlaboratory Evaluation of Test Methods to Determine TAPPI Repeatability and Reproducibility").

Approved by the Standard Specific Interest Group for this Standard Practice
TAPPI

1.3 Both nonstatistical and statistical rules for dealing with suspect test determinations are given. Basically no test determination should be accepted, no matter how correct the value appears to be, if it is known that a faulty determination has been made, and no test determination should be completely rejected purely on a statistical significance test.

1.4 The statistical tests described in this practice have been selected from a large number that are available. They apply to the simplest kind of experimental data, that is, replicate determinations of some property of a given sample of material.

NOTE 1: This practice applies to replicate *test determinations*, usually on several specimens taken under the same conditions and measured in a brief period of time. A test result, obtained in accordance with a TAPPI Test Method, is usually one or the average of two or more such test determinations (see definitions in TAPPI T 400 "Sampling and Accepting a Single Lot of Paper, Paperboard, Containerboard or Related Product"). This practice allows the examination and possible elimination of suspect test determinations (from sets of 3 to 30 determinations) before the calculation of the final test result.

NOTE 2: This practice may also be applied to suspect *test results* (by substituting the words "test results" for "test determinations" throughout this document), when a laboratory must evaluate a large shipment requiring the determination and calculation of several test results.

1.5 Three categories of suspect determinations are considered:

1.5.1 A single suspect determination;

1.5.2 Two suspect determinations, one the least and the other the greatest in the set of replicate determinations; and

1.5.3 Two suspect determinations, the two largest or the two smallest in the set.

2. Summary of procedure

2.1 First, nonstatistical reasons for rejecting or correcting test determinations are considered. Then the remaining apparent outliers are subjected to statistical tests of significance, and those found to be statistically significant are again examined for nonstatistical causes. Finally, appropriate action is taken as regards the further analysis of the data in the presence of outlying test determinations for which no cause can be determined.

2.2 Only statistical tests relatively simple to calculate are included in this standard practice. One group of tests involving the Dixon criteria avoids calculation of the standard deviation (s) and permits quick judgment. Since s is now easily obtainable on a pocket or desk calculator, other tests are included that do require a calculation of s; these are generally more powerful in detecting outliers, except when the Dixon test avoids secondary outliers included in the other tests.

2.3 The tables reproduced in this recommended practice are limited to 25 or 30 replicate determinations.

2.3.1 Generally, it is not useful to make a larger number of determinations in order to reduce the effect of random errors or sample variability, because the effect of a constant error present in all the determinations will usually become more important than the residual effect of the random errors. However, if more than 25 or 30 replicate determinations have been made, the data may be divided into groups of less than 25 or 30 and each group separately tested for outliers.

2.3.2 The tables are based on an assumed underlying normal (Gaussian) distribution (*1*) of test determinations, and those involving the standard deviation assume that the estimate of the standard deviation must come from the group of replicate determinations being tested. If the distribution is not normal but is of the same general bell shape, the level of risk of an erroneous conclusion (see 4.2.1.3) will change somewhat but not seriously. However, if the distribution is markedly skewed, test determinations on the long-tailed side of the distribution will be disproportionately selected out. If the tail is on the high side, as for folding endurance, a log transformation of the data prior to application of the tests for outliers is suggested in order to induce a more normal distribution of the data.

3. Significance

Even a single aberrant determination among a number of replicate test determinations of a property of a sample of material may lead to an incorrect conclusion about the nature of the material. On the other hand, discarding determinations because they appear to be aberrant, when in fact they are not, can also result in false conclusions. This standard practice provides a procedure for reducing these two dangers.

4. Procedure

4.1 *Nonstatistical considerations*

4.1.1 When it is clearly known that a deviation from the prescribed test method (e.g., a blunder) has taken place, discard the resultant determination even when it appears to agree with the rest of the data. Such a deviation might be the accidental rubbing of the observer's finger against the swinging pendulum during the tearing strength test. However, correct and retain the determination if a reliable procedure for doing so is available, such as correcting a change in temperature.

4.2 *Statistical tests for detecting outliers*

4.2.1 *General procedure*

4.2.1.1 If the number of replicate test determinations in the set to be tested is greater than 25 (or 30), divide them into groups of less than 25 (or 30), depending on the statistical test to be used (see Tables 1-4). Divide the determinations in the order in which they were obtained (i.e., first 25, second 25, etc.), unless trends (drifting, cycling, etc.) are observed, in which case assign the determinations at random to the groups.

4.2.1.2 Designate the number of determinations in the group to be tested by the letter *n*, and arrange the *n* determinations in order of magnitude:

$$x_1, x_2, x_3, \ldots\ldots\ldots\ldots x_n$$

The order of the arrangement may be either ascending or descending and the suspected determination(s) may be either the largest or the smallest, or the extremes, depending on the statistical test to be used, as explained in the following sections.

4.2.1.3 Choose an appropriate significance level or risk (probability) of erroneously selecting a good determination. Unless otherwise required, it is conventional to use the 5% level.

4.2.1.4 The statistical test to be used depends on the number and location of the suspect test determinations. For each category of suspect determinations, the non-Dixon test is generally the more powerful in confirming outliers, but is frequently more sensitive to an error in selecting the category. An incorrect choice of statistical test will result in incorrect conclusions, as illustrated by the examples. Therefore, carefully examine the data (arranged as in 4.2.1.2) for suspect test determinations, and select the appropriate statistical test, as follows:

Category	*Dixon test*	*G statistic*	*w/s statistic*	*Modified Grubbs test*
Single suspect determination	4.2.2	4.2.3		
Two suspect determinations, the least and the greatest	4.2.4		4.2.5	
Two suspect determinations, the two largest or two smallest	4.2.6			4.2.7

4.2.1.5 Calculate the statistic required by the statistical test to be used and compare it with the critical value given in the table for the chosen level of risk and the number *n* of replicate test determinations. If the statistic *exceeds* the critical value, conclude that the suspected determination(s) has (have) been confirmed to be an outlier; except that in 4.2.7, if the statistic is *less than* the critical value, conclude that the suspicion is confirmed.

4.2.2 *The Dixon test for a single suspect determination*

4.2.2.1 With the data arranged as in 4.2.1.2, assume that x_1 is the suspect test determination. Depending on the number of test determinations, calculate the appropriate statistic r_1, r_2, r_3, or r_4 [see Table 1 from Dixon (*2*)]. Note that the statistic compares the distance of this suspect determination from its neighbors with the range of all of the *n* determinations (*n* less than or equal to 7), or with all but one, two, or three of the *n* determinations (*n* equal to or greater than 8).

4.2.2.2 Compare the calculated value of r with the tabulated critical value (Table 1), confirming or rejecting the suspicion of an outlying test determination.

4.2.2.3 *Examples:*

(a) Five test determinations yielded the following values (rearranged in decreasing order of magnitude): 0.1064, 0.1057, 0.1056, 0.1055, and 0.1053. The determination 0.1064 is suspect. Since n = 5, the statistic r_1 applies (Table 1):

$$r_1 = \frac{0.1057 - 0.1064}{0.1053 - 0.1064} = 0.636$$

At the 5% risk level, the critical value for n = 5 is 0.642. The calculated statistic r_1 does not exceed this tabulated critical value, so it is concluded that the determination 0.1064 is not an outlier.

(b) Fourteen test determinations yielded the following values (rearranged in increasing order of magnitude): 0.6, 2.0, 2.0, 2.1, 2.1, 2.1, 2.2, 2.2, 2.2, 2.3, 2.3, 2.3, 3.0 and 4.0. Since n = 14, the statistic r_4 applies (Table 1). To check the inclusion of 0.6:

$$r_4 = \frac{2.0 - 0.6}{2.3 - 0.6} = 0.824$$

At the 5% risk level, the critical value for n = 14 is 0.546. The calculated statistic r_4 exceeds this tabulated critical value, so it is concluded that the determination 0.6 is an outlier. Indeed, it is an outlier even at the 1% risk level.

4.2.3 *The ratio of deviation to standard deviation test for a single suspect determination (G-statistic)*

4.2.3.1 With the data arranged as in 4.2.1.2, assume that x_1 is the suspect determination. Calculate the mean $\bar{x}$, the standard deviation s, and the statistic G, as follows:

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

$$s = \left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}\right]^{1/2} \qquad G = \frac{|\bar{x} - x_1|}{s}$$

The G-statistic is more powerful than the Dixon statistic for the single-outlier case.

4.2.3.2 Compare the calculated value of G with the tabulated critical value [Table 2 from Grubbs (*3*) and Grubbs and Beck (*4*)], confirming or rejecting the suspicion of an outlying test determination.

4.2.3.3 *Examples:*

(a) Using the same example as in 4.2.2.3 (a):

$$\bar{x} = 0.1057$$
$$s = 0.0004183$$

$$G = \frac{|0.1057 - 0.1064|}{0.000183} = 1.673$$

At the 5% risk level, the critical value for n = 5 is 1.672. The calculated statistic G exceeds this tabulated critical value, so it is concluded that the determination 0.1064 is an outlier. It will be noted that this conclusion reverses that of the Dixon test. These results illustrate how borderline cases may be rejected under one test and accepted under another. Since the *G*-statistic is the best one to use for the single-outlier case, it should be used for the final statistical judgment.

(b) Using the same example as in 4.2.2.3 (b):

$$\bar{x} = 2.243$$
$$s = 0.710$$

$$G = \frac{|2.243 - 0.60|}{0.710} = 2.31$$

At the 5% risk level, the critical value for n = 14 is 2.371. The calculated statistic *G* does not exceed this tabulated critical value, so it is concluded that the determination 0.6 is not an outlier. Normally the more powerful *G*-statistic should have shown an outlier if the Dixon test does. However, closer examination of the data shows that the *G*-statistic was improperly applied to the case since the data include more than one suspect determination, the others (3.0 and 4.0) at the other extreme being avoided in this case by the Dixon test.

4.2.4 *The Dixon test applied to two suspect determinations, the least and the greatest*

4.2.4.1 With the data arranged as in 4.2.1.2, assume that x_1 and x_n are the suspect determinations.

4.2.4.2 If n is greater than seven, the Dixon test excludes the opposite extreme so that each extreme may be separately tested following the procedure of 4.2.2.

4.2.4.3 If n is less than or equal to seven, examine the data to see which of the two extremes is farthest from its neighbor. Temporarily omit this extreme from the analysis, reduce n by one, and apply the procedure of 4.2.2 to the remaining determinations. If the second extreme is thereby confirmed as an outlier, the first extreme may also be accepted as an outlier without further testing. If the second extreme is not confirmed as an outlier, the first extreme must be tested with all of the data following the procedure of 4.2.2.

4.2.4.4 *Examples:*

(a) Seven test determinations yielded the following values: 3.10, 4.25, 4.37, 4.56, 4.68, 4.98, and 5.92. The determinations 3.10 and 5.92 are suspect. Determination 3.10 is farthest from its neighbor. In the remaining six determinations, determination 5.92 is suspect. Since n = 6, the statistic r_1 may be applied (Table 1):

$$r_1 = \frac{5.92 - 4.98}{5.92 - 4.25} = 0.563$$

At the 5% risk level, the critical value for n = 6 is 0.560. The calculated statistic r_1 exceeds this tabulated critical value, so it is concluded that the determination 5.92 is an outlier. The more extreme determination 3.10 may therefore be classified as an outlier without further testing. It should be noted, however, that the statistical test did not use all of the determinations and therefore conclusions drawn therefrom need careful interpretation.

(b) Seven test determinations yielded the following values: 3.60, 4.75, 4.87, 5.06, 5.18, 5.48, and 6.01. Determinations 3.60 and 6.01 are suspect, with 3.60 being farthest from its neighbor. In the remaining six determinations, 6.01 is suspect.

$$r_1 = \frac{6.01 - 5.48}{6.01 - 4.75} = 0.421$$

At the 5% risk level, the critical value for n = 6 is 0.560. The calculated statistic does not exceed this tabulated critical value, so it is concluded that the determination 6.01 is not an outlier. In the complete set of seven determinations, determination 3.60 was the most suspect.

$$r_1 = \frac{3.60 - 4.75}{3.60 - 6.01} = 0.477$$

At the 5% risk level, the critical value for n = 7 is 0.507. The calculated statistic r_1 does not exceed this tabulated critical value, so it is concluded that the determination 3.60 is also not an outlier.

4.2.5 *The ratio of the range to the standard deviation test for two suspect determinations, the least and the greatest (w/s - statistic)*

4.2.5.1 With the data arranged as in 4.2.1.2, assume that x_1 and x_n are the suspect determinations. Calculate the range w = x_n - x_l, the mean $\bar{x}$, the standard deviation *s* (as in 4.2.3.1), and the statistic *w/s*. The *w/s*-statistic is more powerful than the Dixon or *G*-statistic for confirming an outlier at one end of the range in the presence of another suspect outlier at the other end.

4.2.5.2 Compare the calculated value of *w/s* with the tabulated critical value [Table 3 from David *et al*. (*5*)], confirming or rejecting the suspicion that one or both suspect values are outliers.

4.2.5.3 If it is confirmed that one or both suspect values x_1 and x_n are outliers and if x_1 and x_n are about equally distant from the mean $\bar{x}$, conclude that both are outliers. Otherwise accept the one farthest from the mean as an outlier and test the other against the remaining determinations, using the Dixon test (4.2.2) or the *G*-statistic (4.2.3) for a single suspect determination.

4.2.5.4 *Examples:*

(a) Using the same example as in 4.2.4.4 (a):

$$w = x_n - x_1 = 5.92 - 3.10 = 2.82$$
$$\bar{x} = 4.551$$
$$s = 0.8469$$
$$w/s = 3.33$$

At the 5% risk level, the critical value for n = 7 is 3.22. The calculated statistic *w/s* exceeds this tabulated critical value, so it is concluded that one or both of the determinations 3.10 and 5.92 are outliers. Since the two determinations are approximately equal distance from the mean (4.55 - 3.10 = 1.45 and 5.92 - 4.55 = 1.37), it is concluded that both are outliers.

(b) Using the same example as in 4.2.4.4 (b):

$$w = x_n - x_1 = 6.01 - 3.60 = 2.41$$
$$\bar{x} = 4.993$$
$$s = 0.7445$$
$$w/s = 3.24$$

At the 5% risk level, the critical value for n = 7 is 3.22. The calculated statistic *w/s* exceeds this tabulated critical value, so it is concluded that one or both of the determinations 3.60 and 6.01 are outliers. The determination 3.60, which is farthest from the mean, is confirmed as an outlier. (This reverses the conclusion from the Dixon test, where the determination 3.60 was not found to be an outlier because of the presence of the second suspect determination 6.01.) Omitting the outlier 3.60 and testing the determination 6.01, which is much closer to the mean, against the remaining determinations, it is found on both the Dixon test (4.2.2) and *G*-statistic (4.2.3) the 6.01 is not an outlier.

4.2.6 *The Dixon test applied to two suspect determinations, the two largest or the two smallest.*

4.2.6.1 With the data arranged as in 4.2.1.2, assume that x_1 and x_2 are the suspect determinations. Temporarily omit the more extreme determination x_1 from the analysis, reducing n by one, and apply the procedure 4.2.2 to the remaining determinations. If this lesser extreme determination is thereby confirmed as an outlier, the more extreme determination may also be accepted as an outlier. If this lesser extreme determination is not confirmed an outlier, the greater extreme must be tested with all of the data following the procedure of 4.2.2.

4.2.6.2 *Examples:*

(a) Ten test determinations yielded the following values: 1.00, 1.20, 2.02, 2.21, 2.57, 2.71, 2.92, 3.03, 3.09, and 3.11. The determinations 1.00 and 1.20 are suspect. Omitting the more extreme determination 1.00 and testing the other suspect determination 1.20, now with n = 9 (statistic r_2 of Table 1 applies):

$$r = \frac{2.02 - 1.20}{3.09 - 1.20} = 0.434$$

At the 5% risk level, the critical value for n = 9 is 0.512. The calculated statistic r_2 does not exceed this tabulated critical value, so it is concluded that the determination 1.20, is not an outlier. Following the procedure of 4.2.2 and using all of the data, it is found that the suspect determination 1.00 is also not an outlier.

(b) Using the same example as in 4.2.2.3 (b), consider the suspect determinations to be 3.0 and 4.0. Omitting the more extreme determination 4.0 and testing the other suspect determination 3.0, now with n = 13:

$$r_3 = \frac{3.0 - 2.3}{3.0 - 2.0} = 0.700$$

At the 5% risk level, the critical value for n = 13 is 0.521. The calculated statistic r_3 exceeds this tabulated value, so it is concluded that the determination 3.0 is an outlier. Indeed it is an outlier even at the 1% level. Also, the more extreme determination 4.0 may be classified as an outlier without further testing. It should be noted however that the statistical test did not use all of the determinations, and conclusions drawn therefrom need careful interpretation.

4.2.7 *The ratio of the standard deviations test for two suspect determinations, the two largest or the two smallest (modified Grubbs test)*

4.2.7.1 With the data arranged as in 4.2.1.2, assume that x_1 and x_2 are the suspect determinations. Calculate the standard deviation s of all of the determinations, the standard deviation s_{l2} omitting the suspect determinations, and the statistic s_{12}/s.

4.2.7.2 Compare the calculated value of s_{12}/s with the tabulated critical value [Table 4 derived from Grubbs and Beck (*4*) and Grubbs (*6*)], confirming or rejecting the suspicion that both suspect values are outliers. In this test, the calculated value is significant if it is *less* than the critical value.

4.2.7.3 If the suspicion that both extreme values are outliers is rejected, then test the more extreme value using the Dixon test (4.2.2) or the G-statistic (4.2.3).

4.2.7.4 *Examples:*

(a) Using the same example as in 4.2.6.2 (a):

$s = 0.771$
$s_{12} = 0.413$
$s_{12}/s = 0.536$

At the 5% risk level, the critical value for n = 10 is 0.544. The calculated statistic s_{12}/s is less than the tabulated critical value, so it is concluded that both the determinations 1.00 and 1.20 are outliers. It will be noted that this conclusion reverses that of the Dixon test in 4.2.6.2 (a). The s_{12}/s test is the best one to use for the two-outlier test when both suspect determinations are on the same side of the mean. Hence the conclusion reached by the s_{12}/s test should be accepted.

(b) Use the same example as in 4.2.6.2 (b) which is the same as in 4.2.2.3 (b) and considering suspect the two extreme determinations 3.0 and 4.0:

$s = 0.710$
$s_{12} = 0.464$
$s_{12}/s = 0.653$

At the 5% risk level the critical value for n = 14 is 0.649. The calculated statistic s_{12}/s is not less than this tabulated value, so it is concluded that the determination 3.0 is not an outlier and the determination 4.0 should be tested separately. Normally the more powerful s_{12}/s statistic should have shown two outliers if the Dixon test does. However, closer examination shows the presence of another extreme determination 0.6 which is avoided by the Dixon test.

4.2.8 *Other tests applicable to one, two, or more suspect determinations*

4.2.8.1 An excellent summary of other tests for evaluating suspect determinations is given by Grubbs (*3*). In addition to covering the cases given here for one or two suspect determinations and where the standard deviation is estimated from the group of replicates being tested, Grubbs covers the case of more than two suspect determinations and also tests where the standard deviation is known or obtained from independent estimates.

4.3 *Rechecking nonstatistical causes*

4.3.1 For those test determinations confirmed to be "outliers" by one of the procedures of 4.2, attempt to determine the cause or causes. For example, examine the test specimens if still available on which the outlying test determinations were obtained. Check the laboratory notebooks or original data records to see if some step in the test procedure was poorly carried out. However, as emphasized by Kruskal (*7*), the great danger here is that it is easy after the fact to bias one's memory or approach.

4.3.2 If reasons are found for the outlying test determinations, then scrutinize the other determinations to assure that the same "reasons" are not present in them. If it is confirmed that the reasons do apply only to the outlying test determinations and that such determinations are not likely to be typical of the material discard them.

4.3.3 If reasons are not found for the outlying test determinations, complete the analysis both with and without them. If the broad conclusions of the two analyses are quite different, view any conclusions with very great caution [Kruskal (*7*)]. Judge the question of what is "quite different" on the goal of the experiment.

4.3.3.1 If the goal of the experiment is that of product evaluation (for example, estimating the percent of *A* in a mixture), retain or omit the apparently aberrant test determination based on the experimenter's judgment as to the technical or economic importance of the difference found in 4.3.3. If the purpose of the product evaluation is the acceptance or rejection of a shipment, the question of what should be done with unexplained outliers should be covered by the procurement agreement, particularly for those outliers the inclusion or exclusion of which will determine the acceptance or rejection of the shipment. One procedure might be to require additional sampling and retesting.

5. Report

Report the outlier test used, the outlier or outliers, and what use has been made thereof in the subsequent analysis of the data.

6. Keywords

Test Methods, Testing, Anomalies, Errors

7. Additional information

7.1 Effective date of issue: To Be Assigned

7.2 This revision deletes reference to suspect (outlying) test determinations in interlaboratory studies for evaluating the precision of test methods. This is now covered by T 1200. Editorial changes and corrections were made in 2004, and more editorial changes were made in 2014.

7.3 Related methods: ASTM E178, "Dealing with Outlying Observations," American Society for Testing & Materials, Philadelphia, PA.

References

1. Mandel, J., *The Statistical Analysis of Experimental Data*, Wiley, 1964; or any standard test on statistics.
2. Dixon, W. J., "Processing Data for Outliers," *Biometrics* **9**(1): 74 (1953).
3. Grubbs, F. E., "Procedures for Detecting Outlying Observations in Samples," *Technometrics* **11** (1): 1 (1969).
4. Grubbs, F. E., and Beck, G., "Extensions of Sample and Percentage Points for Significance Tests of Outlying Observations," *Technometrics* **14** (4): 847 (1972).
5. David, H. A., Hartley, H. O., and Pearson, E. X., "The Distribution of the Ratio in a Single Sample of Range to Standard Deviation," *Biometrika* **41:** 482 (1954).
6 Grubbs, F. E., "Sample Criteria for Testing Outlying Observations," *Annals of Mathematical Statistic* **21:** 27 (1950).
7. Kruskal, W. H., "Some Remarks on Wild Observations," *Technometrics* **2:** 1 (1960).

Table 1. Dixon criteria for testing a single extreme determination.

Number of determinations (n)	*Risk* 5%	*Risk* 1%	*Statistic computed*
3	0.941	0.988	$r_1 = \frac{x_2 - x_1}{x_n - x_1}$
4	0.765	0.889	
5	0.642	0.780	
6	0.560	0.698	
7	0.507	0.637	For outlier x_1
8	0.554	0.683	$r_2 = \frac{x_2 - x_1}{x_{n-1} - x_1}$
9	0.512	0.635	
10	0.477	0.597	For outlier x_1 (Avoids x_n)
11	0.576	0.679	$r_3 = \frac{x_3 - x_1}{x_{n-1} - x_1}$
12	0.546	0.642	
13	0.521	0.615	For outlier x_1 (Avoids x_2 and x_n)
14	0.546	0.641	$r_4 = \frac{x_3 - x_1}{x_{n-2} - x_1}$
15	0.525	0.616	
16	0.507	0.595	
17	0.490	0.577	For outlier x_1 (Avoids x_2, x_n, and x_{n-1})
18	0.475	0.561	
19	0.462	0.547	
20	0.450	0.535	
21	0440	0.524	
22	0.430	0.514	
23	0.421	0.505	
24	0.413	0.497	
25	0.406	0.489	

Table 2. Critical values of the *G*-statistic for testing a single extreme determination

Number of determinations (*n*)	Risk 5%	Risk 1%	Statistic computed
3	1.153	1.155	
4	1.463	1.492	$G = \frac{\lvert \bar{x} - x_1 \rvert}{s}$
5	1.672	1.749	
6	1.822	1.944	
7	1.938	2.097	where
8	2.032	2.221	
9	2.110	2.323	
10	2.176	2.410	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
11	2.234	2.485	
12	2.285	2.550	
13	2.331	2.607	
14	2.371	2.659	
15	2.409	2.705	
16	2.443	2.747	$s = \left[\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - l} \right]^{½}$
17	2.475	2.785	
18	2.504	2.821	
19	2.532	2.854	
20	2.557	2.884	
21	2.580	2.912	
22	2.603	2.939	
23	2.624	2.963	
24	2.644	2.987	
25	2.663	3.009	

NOTE 3: Grubbs called this the *T*-statistic; it is called the *G*-statistic here (for "Grubbs") in order to avoid possible confusion with Student's *t* test.

Table 3. Critical values for *w/s* (ratio of range to standard deviation) for simultaneously testing two extreme determinations, the least and the greatest

Number of determinations (n)	Risk 5%	Risk 1%	Statistic computed
3	2.00	2.00	$\frac{w}{s} = \frac{x_n - x_1}{s}$
4	2.43	2.44	
5	2.75	2.80	
6	3.01	3.10	where
7	3.22	3.34	
8	3.40	3.54	
9	3.55	3.72	
10	3.68	3.88	$s = \left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-l}\right]^{½}$
11	3.80	4.01	
12	3.91	4.13	
13	4.00	4.24	
14	4.09	4.34	
15	4.17	4.43	
16	4.24	4.51	
17	4.31	4.59	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
18	4.38	4.66	
19	4.43	4.73	
20	4.49	4.76	
30	4.89	5.25	

Table 4. Critical values for s_{12}/s for simultaneously testing two extreme determinations, the two largest or the two smallest

Number of determinations (n)	*Risk* 5%	*Risk* 1%
5	0.191	0.084
6	0.307	0.176
7	0.391	0.257
8	0.455	0.324
9	0.505	0.380
10	0.544	0.426
11	0.577	0.466
12	0.605	0.500
13	0.629	0.529
14	0.649	0.555
15	0.667	0.578
16	0.683	0.598
17	0.698	0.616
18	0.711	0.633
19	0.722	0.647
20	0.733	0.661
21	0.742	0.673
22	0.751	0.685
23	0.759	0.696
24	0.767	0.705
25	0.774	0.715

Statistic computed

$$\frac{s_{12}}{s}$$

$$s_{12} = \left[\frac{\sum_{i=3}^{n} (x_i - \bar{x}_{12})^2}{n-3} \right]^{1/2}$$

$$\bar{x}_{12} = \frac{1}{n-2} \sum_{i=3}^{n} x_i$$

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

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

Your comments and suggestions on this procedure are earnestly requested and should be sent to the TAPPI Standards Department.

■