



TAPPI/ANSI T 1205 sp-14

TENTATIVE STANDARD – 1963
RECOMMENDED PRACTICE – 1982
REVISED – 1987
REVISED – 1992
REVISED – 2000
REVISED – 2005
REAFFIRMED – 2009
REVISED – 2014
©2014 TAPPI

The information and data contained in this document were prepared by a technical committee of the Association. The committee and the Association assume no liability or responsibility in connection with the use of such information or data, including but not limited to any liability under patent, copyright, or trade secret laws. The user is responsible for determining that this document is the most recent edition published.

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 Material 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

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”).

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.

Approved by the Standard Specific Interest Group for this Standard Practice



Documents
Document Sharing Hub

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:

$$x_1, x_2, x_3, \dots, x_n$$

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:

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

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

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):

$$\bar{x} = 2.2429$$

$$s = 0.7101$$