Designation: E178 – 21

An American National Standard

Standard Practice for Dealing With Outlying Observations[1]

This standard is issued under the fixed designation E178; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This practice covers outlying observations in samples and how to test the statistical significance of outliers.

1.2 The system of units for this standard is not specified. Dimensional quantities in the standard are presented only as illustrations of calculation methods. The examples are not binding on products or test methods treated.

1.3 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety, health, and environmental practices and determine the applicability of regulatory limitations prior to use.*

1.4 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]
E456 Terminology Relating to Quality and Statistics
E2586 Practice for Calculating and Using Basic Statistics

3. Terminology

3.1 *Definitions*—Unless otherwise noted in this standard, all terms relating to quality and statistics are defined in Terminology E456.

3.1.1 *null hypothesis, H_0, n*—a statement about a parameter of a probability distribution or about the type of probability distribution, tentatively regarded as true until rejected using a statistical hypothesis test. **E2586**

3.1.2 *order statistic $x_{(k)}$, n*—value of the kth observed value in a sample after sorting by order of magnitude. **E2586**

3.1.2.1 *Discussion*—In this practice, x_k is used to denote order statistics in place of $x_{(k)}$, to simplify the notation.

3.1.3 *outlier*—see **outlying observation**.

3.1.4 *outlying observation, n*—an extreme observation in either direction that appears to deviate markedly in value from other members of the sample in which it appears.

3.1.4.1 *Discussion*—The identification of a value as outlying, and therefore a doubtful observation, is a judgement of the analyst and can be made before any statistical test.

4. Significance and Use

4.1 An outlying observation, or "outlier," is an extreme one in either direction that appears to deviate markedly from other members of the sample in which it occurs.

4.2 Statistical rules test the null hypothesis of no outliers against the alternative of one or more actual outliers. The procedures covered were developed primarily to apply to the simplest kind of experimental data, that is, replicate measurements of some property of a given material or observations in a supposedly random sample.

4.3 A statistical test may 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.

5. Procedure

5.1 In dealing with an outlier, the following alternatives should be considered:

5.1.1 An outlying observation might be the result of gross deviation from prescribed experimental procedure or an error in calculating or recording the numerical value. When the experimenter is clearly aware that a 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 is available, the observation may sometimes be corrected and retained.

5.1.2 An outlying observation might be merely an extreme manifestation of the random variability inherent in the data. If this is true, the value should be retained and processed in the same manner as the other observations in the sample. Transformation of data or using methods of data analysis designed for a non-normal distribution might be appropriate.

---

[1] This practice is under the jurisdiction of ASTM Committee E11 on Quality and Statistics and is the direct responsibility of Subcommittee E11.10 on Sampling / Statistics.

Current edition approved June 1, 2021. Published June 2021. Originally approved in 1961. Last previous edition approved in 2016 as E178 – 16a. DOI: 10.1520/E0178-21.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

5.1.3 Test units that give outlying observations might be of special interest. If this is true, once identified they should be segregated for more detailed study. Outliers may contain important information for a possible root cause analysis and action on the process or procedure.

5.2 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 clearcut decision to be made.

5.2.1 When the experimenter cannot identify abnormal conditions, they should report the discordant values and indicate to what extent they have been used in the analysis of the data.

5.3 Thus, as part of the over-all process of experimentation, the process of screening samples for outlying observations and acting on them is the following:

5.3.1 *Physical Reason Known or Discovered for Outlier(s):*

5.3.1.1 Reject observation(s) and possibly take additional observation(s).

5.3.1.2 Correct observation(s) on physical grounds.

5.3.2 *Physical Reason Unknown—Use Statistical Test:*

5.3.2.1 Reject observation(s) and possibly take additional observation(s).

5.3.2.2 Transform observation(s) to improve fit to a normal distribution.

5.3.2.3 Use estimation appropriate for non-normal distributions.

5.3.2.4 Segregate samples for further study.

6. Basis of Statistical Criteria for Outliers

6.1 In testing outliers, 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 "significance level" or "percentage point" and can be thought of as the risk of erroneously rejecting a good observation. If a real shift or change in the value of an observation arises from nonrandom causes (human error, loss of calibration of instrument, change of measuring instrument, or even change of time of measurements, and so forth), then the observed value of the sample criterion used will exceed the "critical value" based on random-sampling theory. Tables of critical values are usually given for several different significance levels. In particular for this practice, significance levels 10, 5, and 1 % are used.

NOTE 1—In this practice, we will usually illustrate the use of the 5 % 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 correct.

6.2 Almost all criteria for outliers are based on an assumed underlying normal (Gaussian) population or distribution. The null hypothesis that we are testing in every case is that all observations in the sample come from the same normal population. In choosing an appropriate alternative hypothesis (one or more outliers, separated or bunched, on same side or different sides, and so forth) it is useful to plot the data as shown in the dot diagrams of the figures. When the data are not normally or approximately normally distributed, the probabilities associated with these tests will be different. The experimenter is cautioned against interpreting the probabilities too literally.

6.3 Although our primary interest here is that of detecting outlying observations, some of the statistical criteria presented may also be used to test the hypothesis of normality or that the random sample taken 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 normal observations.

6.4 One should distinguish between data to be used to estimate a central value from data to be used to assess variability. When the purpose is to estimate a standard deviation, it might be seriously underestimated by dropping too many "outlying" observations.

7. Recommended Criteria for Single Samples

7.1 *Criterion for a Single Outlier*—Sort the n observations in order of increasing magnitude by $x_1 \le x_2 \le x_3 \le \ldots \le x_n$, called *order statistics*. Let the largest value, x_n, be the doubtful value, that is the largest value. The test criterion, T_n, for a single outlier is as follows:

$$T_n = (x_n - \bar{x})/s \tag{1}$$

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:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^{\,2} - n \cdot \bar{x}^{\,2}}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i\right)^2 / n}{n-1}} \tag{3}$$

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

$$T_1 = (\bar{x} - x_1)/s \tag{4}$$

The critical values for either case, for the 1, 5, and 10 % levels of significance, are given in Table 1.

7.1.1 The test criterion T_n can be equated to the Student's t test statistic for equality of means between a population with one observation x_n and another with the remaining observations $x_1, \ldots, x_{n-1}$, and the critical value of T_n for significance