# Testing for Outliers

## Practical and Philosophical Considerations

BY THOMAS J. BZIK

### Q: How can you test for an outlier, and Is there an ASTM standard for this practice?

A: An outlier is an observation or subset of observations that appears to be inconsistent with the remainder of the data (Fig. 1). ASTM E178, Practice for Dealing with Outlying Observations, provides statistical procedures for outlier testing. This standard provides criteria and guidance for single outliers as well as multiple outliers.

In typical outlier testing, outliers are identified by their either being extremely large or extremely small relative to the main body of data. Outlier testing does not focus on identifying corrupt data that falls within the main body of data. This is motivated by both the issue that identification of such corrupt data often has to be done on a nonstatistical basis and that corrupt data within the main body of the data has much less impact on sample statistics than do outliers.

Statistical tests for outliers, such as those found in ASTM E178, make the assumption that the true unknown data distribution is normal (Gaussian) in nature. These tests assume that a unique symmetric unimodal mound-shaped underlying distribution is an appropriate data model for testing for outliers. Additionally, outlier tests may be targeted at detecting only one outlier, up to two outliers, or multiple outliers. Prior to outlier testing, it can be useful to examine one of the most basic assumptions, that of a unimodal distribution. Data that exhibits multimodality indicates that there is often additional substructure that may be useful to account for both in data analysis and subsequent reporting of summary statistics. Outlier testing is, to a great degree, premature in such a situation.

Another situation in which the nature of the underlying distribution should be examined more closely is when the data strongly exhibits non-normality. This often will take the form of a skewed or asymmetric distribution. Outlier testing will yield predictable results in these situations, too high a chance of claiming an outlier on the longer-tailed side and too low a chance of identifying an outlier on the shorter-tailed side. The dilemma the data analyst faces is whether the asymmetry was caused by an outlier or outliers or is simply an inherent property of the measurement process. Some data contexts often exhibit strong data asymmetry such as measurement of trace contaminants. Future versions of E178 may include outlier tests that can handle asymmetry or are nonparametric (not based on using mean and standard deviation) in nature.

Outlier tests can be grouped loosely into three data categories depending on what is already known historically about the standard deviation of the process generating the data. There are three possible states of knowledge: 1) no prior knowledge, 2) some prior knowledge and 3) "known." Different outlier tests are used depending on the state of knowledge, for example, the Grubbs test can be used when testing for a single outlier and there is no prior knowledge of the standard deviation. "Known" can be taken to indicate that the historical estimate of the standard deviation is based on a large number of degrees of freedom. Outlier tests based on 1) will generally underperform those that leverage historical knowledge about standard deviation. This underperformance can be substantial when the current data being tested also has a small sample size itself. Tests based on no prior knowledge, such as the Grubbs test, do not perform well for small *n* while tests for which there is some reasonable prior knowledge can perform reasonably well for small *n*.

Consider three cases for the perceived or known prevalence of outliers: 1) very unlikely, 2) can happen, but relatively unlikely and 3) anything more extreme than 1) and 2). In Case 1, it can still make sense to run an outlier test, but use of a smaller than typical alpha is recommended. In Case 2, one is in the best philosophi-

cal scenario for applying typical outlier tests. In Case 3, outlier testing is unlikely to provide a good fix to the problem, and robust estimation methods are preferred rather than outlier testing. Examples of robust (insensitive to outliers) estimates of central tendency include measures such as the median or trimmed means. The sample median, an estimate of the 50th percentile, will typically be only slightly impacted by the addition of an outlier no matter how extreme the outlier is. Trimmed means avoid the impact of outliers by simply not using a set proportion of the most extreme data in performing the calculation. The potential impact of using robust estimates of central tendency for Cases 1 and 2 versus outlier testing is not developed here.

Effective outlier identification becomes extremely critical and relatively riskier when the statistical estimates of interest involve the tail regions of the distribution. Examples include control limits, tail quantiles and possibly risk assessment. Erroneous removal of a statistically identified outlier or erroneous inclusion of a true outlier has much stronger impact in the tails than with measures of central tendency. In these situations, what seems to be a statistically apparent outlier may be the most informative observation.

A fundamental issue in outlier testing is what to do when an outlier test has flagged a data point as a statistical outlier. Should such points be removed from subsequent statistical analysis? This simple question does not have a simple answer. Consider several scenarios for an outlier identified by an outlier test: 1) cause of outlier identified and the problem(s) that caused the outlier are remediated, 2) cause of outlier identified and the problem(s) that caused the outlier are not remediated and 3) cause of outlier not identified. In Case 1, the outlier should be removed from subsequent analysis. For Case 2, if there is another way to catch the occurrence of such an outlier in the future (e.g., violates a physical limit), then remove it, otherwise, keep it or present two analyses – one each with and without this point. In Case 3, the outlier should typically not be removed from subsequent analysis, or at a minimum complete analyses with and without this point should be presented.

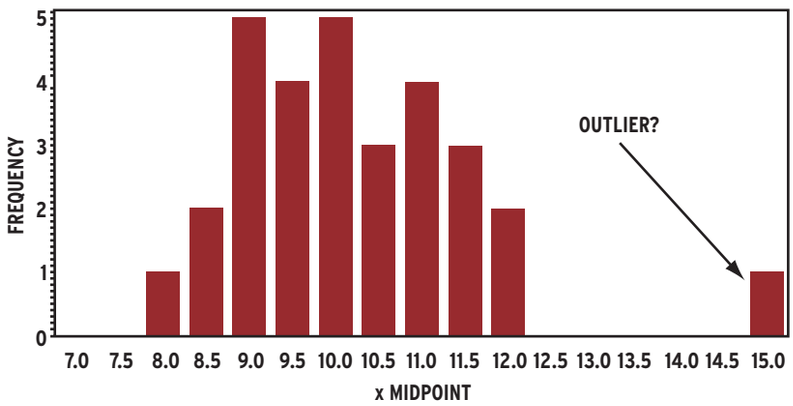


**Figure 1 - Example of a Possible Outlier**

For more, check the Nov.-Dec. 2008 *SN* DataPoints, "Dealing with Outliers," and ASTM E178.

**THOMAS J. BZIK**, *Air Products and Chemicals Inc., Allentown, Pa., is the American Statistical Association representative to ASTM Committee E11 on Quality and Statistics; he serves as vice chair of E11.20 on Test Method Evaluation and Quality Control and as member at large on the E11 executive subcommittee.*

**DEAN V. NEUBAUER**, *Corning Inc., Corning, N.Y., coordinates the DataPoints column; an ASTM International fellow, he is vice chairman of Committee E11 on Quality and Statistics, chairman of Subcommittee E11.30 on Statistical Quality Control and chairman of E11.90.03 on Publications.*

***Statistics play an important role in the ASTM International standards you write, and a panel of experts is ready to answer your questions about how to use statistical principles in ASTM standards. Please send your questions to*** SN ***Editor in Chief Maryann Gorman at mgorman@astm.org or ASTM International, 100 Barr Harbor Drive, P.O. Box C700, West Conshohocken, PA 19428-2959.***