

A Review of Statistical Outlier Methods

Nov 2, 2006

By: [Steven Walfish](#)

Pharmaceutical Technology

Statistical outlier detection has become a popular topic as a result of the US Food and Drug Administration's out of specification (OOS) guidance and increasing emphasis on the OOS procedures of pharmaceutical companies. When a test fails to meet its specifications, the initial response is to conduct a laboratory investigation to seek an assignable cause. As part of that investigation, an analyst looks for an observation in the data that could be classified as an outlier. The FDA guidance "Investigating Out of Specification (OOS) Test Results for Pharmaceutical Production" and the *US Pharmacopeia* are clear that a chemical result cannot be omitted with an outlier test, but that a bioassay can be omitted with an outlier test (1). The two areas specifically prohibited from outlier tests are content uniformity and dissolution testing.



An *outlier* is defined as an observation that "appears" to be inconsistent with other observations in the data set (2). An outlier has a low probability that it originates from the same statistical distribution as the other observations in the data set. On the other hand, an *extreme value* is an observation that might have a low probability of occurrence but cannot be statistically shown to originate from a different distribution than the rest of the data.

Why study outliers

Outliers can provide useful information about the process. An outlier can be created by a shift in the location (mean) or in the scale (variability) of the process. Though an observation in a particular sample might be a candidate as an outlier, the process might have shifted. Sometimes, the spurious result is a gross recording error or a measurement error. Measurement systems should be shown to be capable for the process they measure. Outliers also come from incorrect specifications that are based on the wrong distributional assumptions at the time the specifications are generated.

How to handle outliers

Once an observation is identified—by means of graphical or visual inspection—as a potential outlier, root cause analysis should begin to determine whether an assignable cause can be found for the spurious result. If no root cause can be determined, and a retest can be justified, the potential outlier should be recorded for future evaluation as more data become available. Often, values that seem to be outliers are the right tail of a skewed distribution. When reporting results, it is prudent to report conclusions with and without the suspected outlier in the analysis. Removing data points on the basis of statistical analysis without an assignable cause is not sufficient to throw data away. Robust or nonparametric statistical methods are alternative methods for analysis. Robust statistical methods such as weighted least-squares regression minimize the effect of an outlier observation (3).

There are various approaches to outlier detection depending on the application and number of observations in the data set. Iglewicz and Hoaglin provide a comprehensive text about labeling, accommodation, and identification of outliers (4). Visual inspection alone cannot always identify an outlier and can lead to mislabeling an observation as an outlier. Using a specific function of the observations leads to a superior outlier labeling rule. Because data are used in estimation with classical measures such as the mean being highly sensitive to outliers, statistical methods were developed to accommodate outliers and to reduce their impact on the

analysis. Some of the more commonly used identification methods are discussed in this article.

Box plot

A box plot is a graphical representation of dispersion of the data. The graphic represents the lower quartile (Q_1) and upper quartile (Q_3) along with the median.

The median is the 50th percentile of the data. A lower quartile is the 25th percentile, and the upper quartile is the 75th percentile. The upper and lower fences usually are set a fixed distance from the interquartile range ($Q_3 - Q_1$). Figure 1 shows the upper and lower fences to be set at 1.5 times the interquartile range. Any observation outside these fences is considered a potential outlier. Even when data are not normally distributed, a box plot can be used because it depends on the median and not the mean of the data.

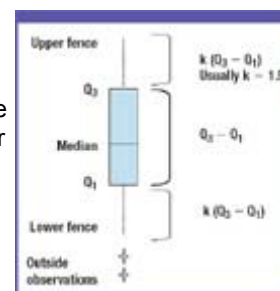


Figure 1: Decomposition of a box plot.

Trimmed means

A trimmed mean is calculated by discarding a certain percentage of the lowest and the highest scores and then computing the mean of the remaining scores. The trimmed mean has the advantage of being relatively resistant to outliers. When outliers are present in the data, trimmed means are robust estimators of the population mean that are relatively insensitive to the outlying values. After viewing the box plot, a potential outlier might be identified. If the upper and lower 5% of the data are removed, then it creates a 10% trimmed mean. An example of trimmed means is the recent change in the Olympic scoring system for ice skating, in which the highest and lowest scores are eliminated, and the mean of the remaining scores is used to assess skaters' scores. If a trimmed mean is presented, the untrimmed mean should be presented for comparison.

Figure 1

Extreme studentized deviate

The extreme studentized deviate (ESD) test is quite good at identifying a single outlier in a normal sample. The maximum deviation from the mean:

$$\frac{|x_i - \bar{x}|}{s}$$

is calculated and compared with a tabled value (see Table I). If the maximum deviation is greater than the tabled value, then the observation is removed, and the procedure is repeated. If no observation exceeds the tabled value, then we cannot claim there is a statistical outlier. A downfall of this method is that it requires the assumption of a normal data distribution. Usually, this assumption holds true as the sample size gets larger, though a formal test such as the Andersen–Darling method can be used to test the assumption (5). This approach can be generalized to investigate multiple outliers simultaneously. Table I is an example of 10 observations (raw data). Based on Table II, the critical value for $N = 10$ at an α level of 0.05 is 2.29. Therefore, the data value 16.3 is an outlier because it corresponds to a studentized deviation of 2.49, which exceeds the 2.29 critical value.

Table I: Example of extreme studentized deviate test.

	Raw data	Studentized deviation
	7.0	0.12
	6.4	0.27
	8.7	0.37
	5.8	0.44
	6.9	0.54
	5.4	0.55
	5.3	0.58
	4.6	0.77
	4.1	0.91
	10.3	2.49
Mean	7.37	
Standard deviation	1.58	

Table I: Example of extreme studentized deviate test.

Table II: Critical values for the extreme studentized deviate test (Reference 4).

n	$\alpha = 0.05$	$\alpha = 0.01$
10	2.24	2.46
11	2.25	2.50
12	2.25	2.54
13	2.26	2.57
14	2.26	2.60
15	2.26	2.63
16	2.27	2.66
17	2.27	2.69
18	2.27	2.71
19	2.28	2.73
20	2.28	2.75
25	2.29	2.79
30	2.29	2.81
40	2.30	2.84
45	2.30	2.85
50	2.30	2.86

Table II: Critical values for the extreme studentized deviate test (Reference 4).

Dixon-type tests

Dixon-type tests are based on the ratio of the ranges. These tests are flexible enough to allow for specific observations to be tested. They also perform well with small sample sizes. Because they are based on ordered statistics, there is no need to assume normality of the data. Depending on the number of suspected outliers, different ratios are used to identify potential outliers. The first class of ratios, r_{10} , is used when the suspected outlier is the largest or smallest observation. The second set of ratios, r_{11} , is used when the potential observation is the second smallest or second largest. Situations like these

arise because of masking. Masking occurs when several observations are close together, but the group of observations is still outlying from the rest of the data. Masking is a common phenomenon especially for bimodal data (*i.e.*, data from two distributions). There are additional sets of ratios depending on how many masked points are excluded (6). The following equations are the r_{10} and r_{11} ratios:

Testing the largest observation as an outlier:

$$r_{10} = \frac{x_n - x_{n-1}}{x_n - x_1}$$

Testing the smallest observation as an outlier:

$$r_{10} = \frac{x_2 - x_1}{x_n - x_1}$$

Testing the largest observation as an outlier avoiding the smallest observation:

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

Testing the smallest observation as an outlier avoiding the largest observation:

$$r_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1}$$

If the distance between the potential outlier to its nearest neighbor is large enough, it would be considered an outlier. Table III shows the critical values for r_{10} and r_{11} ratios.

Using the data set in Table I, the Dixon-type test can be used to determine whether 16.3 is a potential outlier. For r_{10} , the test statistic is $(16.3 - 9.3)/(16.3 - 4.1)$ which is equal to 0.574 and is greater than the tabled value of 0.412. Therefore, for a sample size of 10, 16.3 is a statistical outlier.

Outliers in regression

Regression analysis or least-squares estimation is a statistical technique to estimate a linear relationship between two variables. This technique is highly sensitive to outliers and influential observations. Outliers in regression can overstate the coefficient of determination (R^2), give erroneous values for the slope and intercept, and in many cases lead to false conclusions about the model. Outliers in regression usually are detected with graphical methods such as residual plots including deleted residuals.

A common statistical test is Cook's distance measure, which provides an overall measure of the impact of an observation on the estimated regression coefficient (7). Just because the residual plot or Cook's distance measure test identifies an observation as an outlier does not mean one should automatically eliminate the point. One should fit the regression equation with and without the suspect point and compare the coefficients of the model and review the mean-squared error and R^2 from the two models.

Summary

Various methods for detecting outliers are used several times during an analysis. The first step in outlier detection is to plot the data using methods such as histograms, scatter plots, or a box plot. The extreme studentized deviate test is an excellent test for data sets that have more than 10 observations from a normally distributed sample. For less than 10 observations, the Dixon test is a good method and does not require distributional assumptions. When performing regression analysis, always review residual plots to ensure no outliers are affecting the model coefficients and inflating the R^2 value. Table IV summarizes these methods and their ease of use.

Steven Walfish is the president of Statistical Outsourcing Services, 403 King Farm Blvd., Suite 201, Rockville, MD 20850, tel. 301.325.3129, fax 301.330.2143, steven@statisticaloutsourcingservices.com, www.statisticaloutsourcingservices.com.

Submitted: June 12, 2006. Accepted: Aug. 14, 2006.

Keywords: analytical testing, manufacturing, statistics

References

1. *USP 29-NF 24* (US Pharmacopeial Convention, Rockville, MD, 2006).
2. V. Barnett and T. Lewis, *Outliers in Statistical Data* (John Wiley & Sons, 2d ed., New York, NY, 1985).

Table III: Critical value for Dixon tests ($\alpha = 0.05$) (Reference 6).

Total number of observations	r_{10}	r_{11}
3	0.941	
4	0.765	0.955
5	0.642	0.807
6	0.560	0.689
7	0.507	0.610
8	0.468	0.554
9	0.437	0.512
10	0.412	0.477

Table III: Critical value for Dixon tests ($\alpha = 0.05$) (Reference 6).

Table IV: Comparison of methods.

Method	Sample size	Number of outliers	Ease of use
Box plot	≥ 25	Multiple	Most statistical software has the option to produce the graph.
Extreme studentized test	> 10	Multiple	Moderately easy to use, though not usually found in statistical software.
Dixon	< 10	1 or 2	Easily calculated, though not usually found in statistical software.

Table IV: Comparison of methods.

3. P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection* (John Wiley & Sons, New York, NY, 1987).
4. B. Iglewicz and D.C. Hoaglin, *How to Detect and Handle Outliers* (American Society for Quality Control, Milwaukee, WI, 1993).
5. M.A. Stephens, "Tests Based on EDF Statistics," in *Goodness-of-Fit Techniques*, R.B. D'Agostino and M.A. Stephens, Eds. (Marcel Dekker, New York, NY, 1986).
6. *CRC Standard Probability and Statistics*, W.H. Beyer, Ed. (CRC Press, Boca Raton, FL).
7. J. Neter *et al.*, *Applied Linear Statistical Models* (Irwin, Chicago, IL, 1985).