

Accuracy, Precision, Mean and Standard Deviation

ICP Operations Guide: Part 14 By Paul Gaines, Ph.D.

[Overview](#)
[Accuracy](#)
[Precision](#)
[Mean](#)
[Standard Deviation](#)
[Further Reading](#)
[View as one page](#)

Overview

There are certain basic concepts in analytical chemistry that are helpful to the analyst when treating analytical data. This section will address accuracy, precision, mean, and deviation as related to chemical measurements in the general field of analytical chemistry.

Accuracy

In analytical chemistry, the term 'accuracy' is used in relation to a chemical measurement. The International Vocabulary of Basic and General Terms in Metrology (VIM) defines *accuracy of measurement* as... "closeness of the agreement between the result of a measurement and a true value." The VIM reminds us that accuracy is a "qualitative concept" and that a *true value* is indeterminate by nature. In theory, a *true value* is that value that would be obtained by a perfect measurement. Since there is no perfect measurement in analytical chemistry, we can never know the *true value*.

Our inability to perform perfect measurements and thereby determine *true values* does not mean that we have to give up the concept of accuracy. However, we must add the reality of *error* to our understanding. For example, let's call a measurement we make X_1 and give the symbol μ for the *true value*. We can then define the *error* in relation to the *true value* and the measured value according to the following equation:

$$\text{error} = X_1 - \mu \quad (14.1)$$

We often speak of accuracy in qualitative terms such as "good," "expected," "poor," and so on. However, we have the ability to make quantitative measurements. We therefore have the ability to make quantitative estimates of the error of a given measurement. Since we can estimate the error, we can also estimate the accuracy of a measurement. In addition, we can define error as the difference between the measured result and the *true value* as shown in equation 14.1 above. However, we cannot use equation 14.1 to calculate the *exact error* because we can never determine the *true value*. We can, however, estimate the error with the introduction of the '*conventional true value*' which is more appropriately called either the assigned value, the best estimate of a true value, the conventional value, or the reference value. Therefore, the error can be estimated using equation 14.1 and the *conventional true value*.

Errors in analytical chemistry are classified as *systematic* (determinate) and *random* (indeterminate). The VIM definitions of *error*, *systematic error*, and *random error* follow:

- **Error** – the result of a measurement minus a true value of the measurand.
- **Systematic Error** – the mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions, minus a true value of the measurand.
- **Random Error** – the result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions.

A **systematic error** is caused by a defect in the analytical method or by an improperly functioning instrument or analyst. A procedure that suffers from a systematic error is always going to give a mean value that is different from the true value. The term 'bias' is sometimes used when defining and describing a systematic error. The measured value is described as being biased high or low when a systematic error is present and the calculated uncertainty of the measured value is sufficiently small to see a definite difference when a comparison of the measured value to the *conventional true value* is made.

Some analysts prefer the term 'determinate' instead of systematic because it is more descriptive in stating that this type of error can be determined. A systematic error can be estimated, but it cannot be known with certainty because the true value cannot be known. Systematic errors can therefore be avoided, i.e., they are determinate. Sources of systematic errors include spectral interferences, chemical standards, volumetric ware, and analytical balances where an improper calibration or use will result in a systematic error, i.e., a dirty glass pipette will always deliver less than the intended volume of liquid and a chemical standard that has an assigned value that is different from the *true value* will always bias the measurements either high or low and so on. The possibilities seem to be endless.

Random errors are unavoidable. They are unavoidable due to the fact that every physical measurement has limitation, i.e., some uncertainty. Using the utmost of care, the analyst can only obtain a weight to the uncertainty of the balance or deliver a volume to the uncertainty of the glass pipette. For example, most four-place analytical balances are accurate to ± 0.0001 grams. Therefore, with care, an analyst can measure a 1.0000 gram weight (true value) to an accuracy of ± 0.0001 grams where a value of 1.0001 to 0.999 grams would be within the random error of measurement. If the analyst touches the weight with their finger and obtains a weight of 1.0005 grams, the total error = $1.0005 - 1.0000 = 0.0005$ grams and the random and systematic errors could be estimated to be 0.0001 and 0.0004 grams respectively. Note that the systematic error could be as great as 0.0006 grams, taking into account the uncertainty of the measurement.

A truly random error is just as likely to be positive as negative, making the average of several measurements more reliable than any single measurement. Hence, taking several measurements of the 1.0000 gram weight with the added weight of the fingerprint, the analyst would eventually report the weight of the fingerprint as 0.0005 grams where the random error is still 0.0001 grams and the systematic error is 0.0005 grams. However, random errors set a limit upon accuracy no matter how many replicates are made.

Precision

The term precision is used in describing the agreement of a set of results among themselves. Precision is usually expressed in terms of the deviation of a set of results from the arithmetic mean of the set (mean and standard deviation to be discussed later in this section). The student of analytical chemistry is taught – correctly – that good precision does not mean good accuracy. However, it sounds reasonable to assume otherwise.

Why doesn't good precision mean we have good accuracy? We know from our discussion of error that there are systematic and random errors. We also know that the total error is the sum of the systematic error and random error. Since truly random error is just as likely to be negative as positive, we can reason that a measurement that has only random error is accurate to within the precision of measurement and the more precise the measurement, the better idea we have of the *true value*, i.e., there is no bias in the data. In the case of random error only, good precision indicates good accuracy.

Now let's add the possibility of systematic error. We know that systematic error will produce a bias in the data from the *true value*. This bias will be negative or

positive depending upon the type and there may be several systematic errors at work. Many systematic errors can be repeated to a high degree of precision. Therefore, it follows that systematic errors prevent us from making the conclusion that good precision means good accuracy. When we go about the task of determining the accuracy of a method, we are focusing upon the identification and elimination of systematic errors. Don't be misled by the statement that 'good precision is an indication of good accuracy.' Too many systematic errors can be repeated to a high degree of precision for this statement to be true.

The VIM uses the terms 'repeatability' and 'reproducibility' instead of the more general term 'precision.' The following definitions and notes are taken directly from the VIM:

- *Repeatability* (of results of measurements) – the closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement.

Additional Notes:

1. These conditions are called repeatability conditions.
2. Repeatability conditions include the same measurement procedure, the same observer, the same measuring instrument, used under the same conditions, the same location, and repetition over a short period of time.

- *Reproducibility* (of results of measurement) – the closeness of the agreement between the results of measurements of the same measurand carried out under changed conditions of measurement.

Additional Notes:

1. A valid statement of reproducibility requires specification of the conditions changed.
2. The changed conditions may include principle of measurement, method of measurement, observer, measuring instrument, reference standard, location, conditions of use, and time.

When discussing the precision of measurement data, it is helpful for the analyst to define how the data are collected and to use the term 'repeatability' when applicable. It is equally important to **specify the conditions used** for the collection of 'reproducibility' data.

Mean

The definition of *mean* is, "an average of n numbers computed by adding some function of the numbers and dividing by some function of n." The central tendency of a set of measurement results is typically found by calculating the arithmetic mean (\bar{x}) and less commonly the median or geometric mean. The mean is an estimate of the *true value* as long as there is no systematic error. In the absence of systematic error, the mean approaches the true value (μ) as the number of measurements (n) increases. The frequency distribution of the measurements approximates a bell-shaped curve that is symmetrical around the mean. The arithmetic mean is calculated using the following equation:

$$\bar{x} = (X_1 + X_2 + \dots + X_n) / n \quad (14.2)$$

Typically, insufficient data are collected to determine if the data are evenly distributed. Most analysts rely upon quality control data obtained along with the sample data to indicate the accuracy of the procedural execution, i.e., the absence of systematic error(s). The analysis of at least one QC sample with the unknown sample(s) is strongly recommended.

Even when the QC sample is in control it is still important to inspect the data for outliers. There is a third type of error typically referred to as a 'blunder'. This is an error that is made unintentionally. A blunder does not fall in the systematic or random error categories. It is a mistake that went unnoticed, such as a transcription error or a spilled solution. For limited data sets (n = 3 to 10), the range ($X_n - X_1$), where X_n is the largest value and X_1 is the smallest value, is a good estimate of the precision and a useful value in data inspection. In the situation where a limited data set has a suspicious outlier and the QC sample is in control, the analyst should calculate the range of the data and determine if it is significantly larger than would be expected based upon the QC data. If an explanation cannot be found for an outlier (other than it appears too high or low), there is a convenient test that can be used for the rejection of possible outliers from limited data sets. This is the *Q test*.

The Q test is commonly conducted at the 90% confidence level but the following table (14-3) includes the 96% and 99% levels as well for your convenience. At the 90% confidence level, the analyst can reject a result with 90% confidence that an outlier is significantly different from the other results in the data set. The Q test involves dividing the difference between the outlier and its nearest value in the set by the range, which gives a quotient – **Q**. The range is always calculated by including the outlier, which is automatically the largest or smallest value in the data set. If the quotient is greater than the rejection quotient, **Q_{0,90}**, then the outlier can be rejected.

Table 14.3: The Q Test

n	Q _{0,90}	Q _{0,96}	Q _{0,99}
3	0.94	0.98	0.99
4	0.76	0.85	0.93
5	0.64	0.73	0.82
6	0.56	0.64	0.74
7	0.51	0.59	0.68
8	0.47	0.64	0.53
9	0.44	0.51	0.60
10	0.41	0.48	0.57

Example: This example will test four results in a data set--1004, 1005, 1001, and 981.

- The range is calculated: 1005 – 981 = 24.
- The difference between the questionable result (981) and its nearest neighbor is calculated: 1001 – 981 = 20.
- The quotient is calculated: 20/24 = 0.83.
- The calculated quotient is compared to the Q_{0,90} value of 0.76 for n=4 (from table 14.3 above) and found to be greater.

- The questionable result (981) is rejected.

Standard Deviation

A useful and commonly used measure of precision is the *experimental standard deviation* defined by the VIM as... "for a series of n measurements of the same measurand, the quantity s characterizing the dispersion of the results and given by the formula:

$$s = [\sum (x_i - \bar{x})^2 / (n-1)]^{1/2} \quad (14.4)$$

x_i being the result of the i-th measurement and \bar{x} being the arithmetic mean of the n results considered."

The above definition is for estimating the standard deviation for n values of a sample of a *population* and is always calculated using n-1. The standard deviation of a population is symbolized as σ and is calculated using n. Unless the entire population is examined, σ cannot be known and is estimated from samples randomly selected from it. For example, an analyst may make four measurements upon a given production lot of material (population). The standard deviation of the set (n=4) of measurements would be estimated using (n-1). If this analysis was repeated several times to produce several sample sets (four each) of data, it would be expected that each set of measurements would have a different mean and a different estimate of the standard deviation.

The *experimental standard deviations of the mean* for each set is calculated using the following expression:

$$s / (n)^{1/2} \quad (14.5)$$

Using the above example, where values of 1004, 1005, and 1001 were considered acceptable for the calculation of the mean and the experimental standard deviation the mean would be 1003, the experimental standard deviation would be 2 and the standard deviation of the mean would be 1.

Significant figures will be discussed along with calculation of the uncertainty of measurement in the next part of this series. ■

Further Reading

- [Part 15: Significant Figures and Uncertainty](#)
- [ICP Operations Guide: Table of Contents](#)
- [More Guides and Papers](#)