

Statistical Treatment of Data

Many times during the course of the Chemistry 115 laboratory you will be asked to report an average, relative deviation, and a standard deviation. You may also have to analyze multiple trials to decide whether or not a certain piece of data should be discarded. This section describes these procedures.

Average and Standard Deviation

The **average** or **mean** of the data set, \bar{x} , is defined by:

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

where x_i is the result of the i^{th} measurement, $i = 1, \dots, N$. The standard deviation, σ , measures how closely values are clustered about the mean. The standard deviation for small samples is defined by:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}}$$

The smaller the value of σ , the more closely packed the data are about the mean, and we say that the measurements are **precise**. In contrast, a high **accuracy** of the measurements occurs if the mean is close to the real result (presuming we know that information). It is easy to tell if your measurements are precise, but it is often difficult to tell if they are accurate.

Relative Deviation

The relative average deviation, d , like the standard deviation, is useful to determine how data are clustered about a mean. The advantage of a relative deviation is that it incorporates the relative numerical magnitude of the average. The relative average deviation, d , is calculated in the following way.

1. Report the relative average deviation (ppt) in addition to the standard deviation in all experiments.

Analysis of Poor Data

1. Keep in mind that you also always have the right to discard a piece of data that you are sure is of low quality: that is, when you are aware of a poor collection. However, beware of discarding data that do not meet the 4d test. You may be discarding your most accurate determination!

Other important concepts and procedures

- Normal error curve: Histogram of an infinitely large number of good measurements usually follow a Gaussian distribution
- Confidence limit (95%)
- Linear least squares fit
- Residual sum of squares
- Correlation coefficient

This page titled [Statistical Treatment of Data](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [W. R. Fawcett](#), [John Berg](#), [P. B. Kelley](#), [Carlito B. Lebrilla](#), [Gang-yu Liu](#), [Delmar Larsen](#), [Paul Hrvatin](#), [David Goodin](#), and [Brooke McMahon](#).