Deviation

- Deviation:

 serves as a measure to quantify the disparity between an [observed value](https://en.wikipedia.org/wiki/Observed_value) of a variable and another designated value, frequently the mean of that variable. Deviations with respect to the [sample mean](https://en.wikipedia.org/wiki/Sample_mean) and the [population mean](https://en.wikipedia.org/wiki/Population_mean) (or "[true value](https://en.wikipedia.org/wiki/True_value)") are called [*errors* and *residuals*](https://en.wikipedia.org/wiki/Errors_and_residuals), respectively. The [sign](https://en.wikipedia.org/wiki/Sign_(mathematics)) of the deviation reports the direction of that difference: the deviation is positive when the observed value exceeds the reference value. The [absolute value](https://en.wikipedia.org/wiki/Absolute_value) of the deviation indicates the size or magnitude of the difference. In a given [sample](https://en.wikipedia.org/wiki/Sample_(statistics)), there are as many deviations as [sample points](https://en.wikipedia.org/wiki/Sample_point). [Summary statistics](https://en.wikipedia.org/wiki/Summary_statistics) can be derived from a set of deviations, such as the [*standard deviation*](https://en.wikipedia.org/wiki/Standard_deviation) and the [*mean absolute deviation*](https://en.wikipedia.org/wiki/Mean_absolute_deviation), measures of [dispersion](https://en.wikipedia.org/wiki/Statistical_dispersion), and the [*mean signed deviation*](https://en.wikipedia.org/wiki/Mean_signed_deviation).

* The deviation of each data point is calculated by subtracting the mean of the data set from the individual data point. Mathematically, the deviation d of a data point x in a data set is given by

�=�−����d=x-mean

* This calculation represents the "distance" of a data point from the mean and provides information about how much individual values vary from the average. Positive deviations indicate values above the mean, while negative deviations indicate values below the mean.
* The sum of squared deviations is a key component in the calculation of variance, another measure of the spread or dispersion of a data set. Variance is calculated by averaging the squared deviations. Deviation is a fundamental concept in understanding the distribution and variability of data points in statistical analysis.
* Types

Signed deviations:

A deviation that is a difference between an observed value and the *true value* of a quantity of interest (such as the population mean) is an *error*.

A deviation that is the difference between the observed value and an estimate of the true value (e.g. the sample mean) is a *residual*. These concepts are applicable for data at the [interval](https://en.wikipedia.org/wiki/Level_of_measurement#Interval_scale) and [ratio](https://en.wikipedia.org/wiki/Level_of_measurement#Ratio_scale) levels of measurement

Unsigned or absolute deviation:

Absolute deviation in statistics is a metric that measures the overall difference between individual data points and a central value, typically the mean or median of a dataset. It is determined by taking the absolute value of the difference between each data point and the central value and then averaging these absolute differences.[[4]](https://en.wikipedia.org/wiki/Deviation_(statistics)#cite_note-4) The formula is expressed as follows:

where

*Di* is the absolute deviation,

*xi* is the data element,

*m*(*X*) is the chosen measure of [central tendency](https://en.wikipedia.org/wiki/Central_tendency) of the data set—sometimes the [mean](https://en.wikipedia.org/wiki/Mean) (�¯), but most often the [median](https://en.wikipedia.org/wiki/Median).

* The average absolute deviation (AAD) in statistics is a measure of the dispersion or spread of a set of data points around a central value, usually the mean or median. It is calculated by taking the average of the absolute differences between each data point and the chosen central value. AAD provides a measure of the typical magnitude of deviations from the central value in a dataset, giving insights into the overall variability of the data.
* Least absolute deviation (LAD) is a statistical method used in regression analysis to estimate the coefficients of a linear model. Unlike the more common least squares method, which minimizes the sum of squared vertical distances (residuals) between the observed and predicted values, the LAD method minimizes the sum of the absolute vertical distances.
* In the context of linear regression, if (*x*1,*y*1), (*x*2,*y*2), ... are the data points, and *a* and *b* are the coefficients to be estimated for the linear model:
* the least absolute deviation estimates (*a* and *b*) are obtained by minimizing the sum.