Deviation (statistics)

In mathematics and statistics, **deviation** is a measure of difference between the observed value of a variable and some other value, often that variable's mean. The sign of the deviation reports the direction of that difference (the deviation is positive when the observed value exceeds the reference value). The magnitude of the value indicates the size of the difference.

Types

A deviation that is a difference between an observed value and the *true value* of a quantity of interest (where *true value* denotes the Expected Value, 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; the Expected Value of a sample can be used as an estimate of the Expected Value of the population) is a **residual**. These concepts are applicable for data at the interval and ratio levels of measurement.

Unsigned or absolute deviation

In statistics, the **absolute deviation** of an element of a data set is the absolute difference between that element and a given point. Typically the deviation is reckoned from the central value, being construed as some type of average, most often the median or sometimes the mean of the data set:

$$D_i = |x_i - m(X)|,$$

where

D_i is the absolute deviation,

x_i is the data element,

$m(X)$ is the chosen measure of central tendency of the data set—sometimes the mean ($\overline{x}$), but most often the median.

Measures

Mean signed deviation

For an unbiased estimator, the average of the signed deviations across the entire set of all observations from the unobserved population parameter value averages zero over an arbitrarily large number of samples. However, by construction the average of signed deviations of values from the sample mean value is always zero, though the average signed deviation from another measure of central tendency, such as the sample median, need not be zero.

Dispersion

Statistics of the distribution of deviations are used as measures of statistical dispersion.

- **Standard deviation** is the frequently used measure of dispersion: it uses squared deviations, and has desirable properties, but is not robust.

- **Average absolute deviation,** is the sum of absolute values of the deviations divided by the number of observations.
- **Median absolute deviation** is a robust statistic which uses the median, not the mean, of absolute deviations.
- **Maximum absolute deviation** is a highly non-robust measure, which uses the maximum absolute deviation.

Normalization

Deviations have units of the measurement scale (for instance, meters if measuring lengths). One can nondimensionalize in two ways.

One way is by dividing by a measure of scale (statistical dispersion), most often either the population standard deviation, in standardizing, or the sample standard deviation, in studentizing (e.g., Studentized residual).

One can scale instead by *location,* not *dispersion:* the formula for a **percent deviation** is the observed value minus accepted value divided by the accepted value multiplied by 100%.

See also

- Anomaly (natural sciences)
- Squared deviations
- Deviate (statistics)
- Variance