Skewness

**Skewness:**

is a measure of the asymmetry of the [probability distribution](https://en.wikipedia.org/wiki/Probability_distribution) of a [real](https://en.wikipedia.org/wiki/Real_number)-valued [random variable](https://en.wikipedia.org/wiki/Random_variable) about its mean. The skewness value can be positive, zero, negative, or undefined.

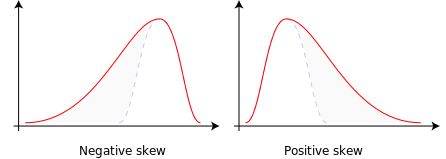
For a [unimodal](https://en.wikipedia.org/wiki/Unimodal) distribution (a distribution with a single peak), negative skew commonly indicates that the *tail* is on the left side of the distribution, and positive skew indicates that the tail is on the right. In cases where one tail is long but the other tail is fat, skewness does not obey a simple rule. For example, a zero value in skewness means that the tails on both sides of the mean balance out overall; this is the case for a symmetric distribution but can also be true for an asymmetric distribution where one tail is long and thin, and the other is short but fat. Thus, the judgement on the symmetry of a given distribution by using only its skewness is risky; the distribution shape must be taken into account.

Introduction

Consider the two distributions in the figure just below. Within each graph, the values on the right side of the distribution taper differently from the values on the left side. These tapering sides are called *tails*, and they provide a visual means to determine which of the two kinds of skewness a distribution has:

*negative skew*: The left tail is longer; the mass of the distribution is concentrated on the right of the figure. The distribution is said to be *left-skewed*, *left-tailed*, or *skewed to the left*, despite the fact that the curve itself appears to be skewed or leaning to the right; *left* instead refers to the left tail being drawn out and, often, the mean being skewed to the left of a typical center of the data. A left-skewed distribution usually appears as a *right-leaning* curve.[[1]](https://en.wikipedia.org/wiki/Skewness#cite_note-cnx.org-1)

*positive skew*: The right tail is longer; the mass of the distribution is concentrated on the left of the figure. The distribution is said to be *right-skewed*, *right-tailed*, or *skewed to the right*, despite the fact that the curve itself appears to be skewed or leaning to the left; *right* instead refers to the right tail being drawn out and, often, the mean being skewed to the right of a typical center of the data. A right-skewed distribution usually appears as a *left-leaning* curve.[[1]](https://en.wikipedia.org/wiki/Skewness#cite_note-cnx.org-1)

[](https://en.wikipedia.org/wiki/File:Negative_and_positive_skew_diagrams_(English).svg)

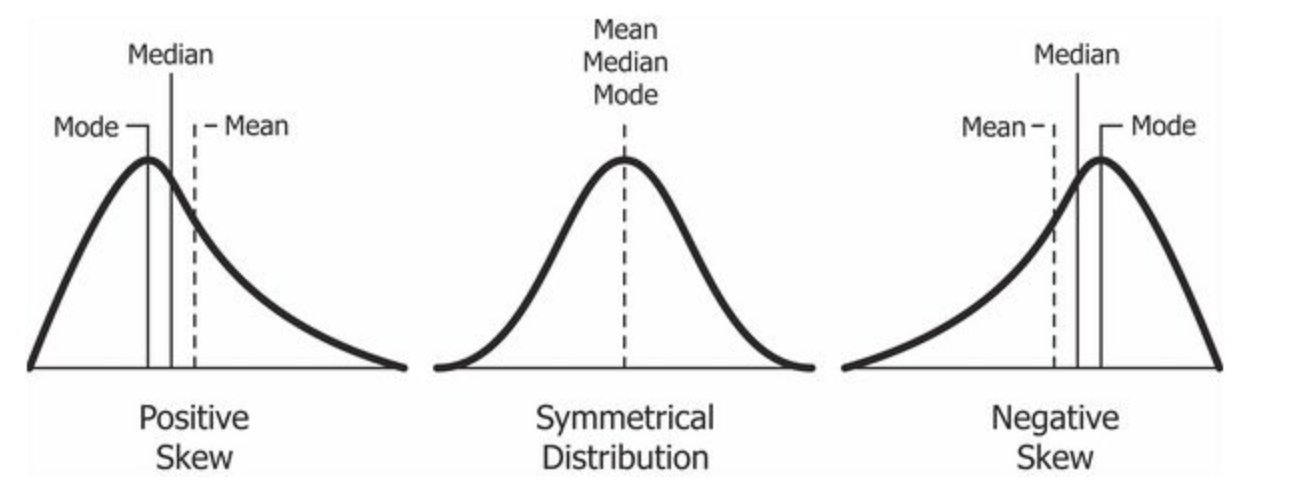
Skewness in a data series may sometimes be observed not only graphically but by simple inspection of the values. For instance, consider the numeric sequence (49, 50, 51), whose values are evenly distributed around a central value of 50. We can transform this sequence into a negatively skewed distribution by adding a value far below the mean, which is probably a negative [outlier](https://en.wikipedia.org/wiki/Outlier), e.g. (40, 49, 50, 51). Therefore, the mean of the sequence becomes 47.5, and the median is 49.5. Based on the formula of [nonparametric skew](https://en.wikipedia.org/wiki/Nonparametric_skew), defined as (�−�)/�, the skew is negative. Similarly, we can make the sequence positively skewed by adding a value far above the mean, which is probably a positive outlier, e.g. (49, 50, 51, 60), where the mean is 52.5, and the median is 50.5.

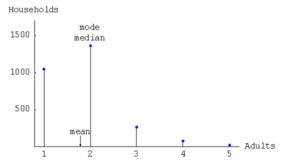
As mentioned earlier, a unimodal distribution with zero value of skewness does not imply that this distribution is symmetric necessarily. However, a symmetric unimodal or multimodal distribution always has zero skewness.

**Relationship of mean and median**

The skewness is not directly related to the relationship between the mean and median: a distribution with negative skew can have its mean greater than or less than the median, and likewise for positive skew.[[](https://en.wikipedia.org/wiki/Skewness#cite_note-von_Hippel_2005-2)

In the older notion of [nonparametric skew](https://en.wikipedia.org/wiki/Nonparametric_skew), defined as (�−�)/�,where �is the [mean](https://en.wikipedia.org/wiki/Mean), �is the [median](https://en.wikipedia.org/wiki/Median), and � is the [standard deviation](https://en.wikipedia.org/wiki/Standard_deviation), the skewness is defined in terms of this relationship: positive/right nonparametric skew means the mean is greater than (to the right of) the median, while negative/left nonparametric skew means the mean is less than (to the left of) the median. However, the modern definition of skewness and the traditional nonparametric definition do not always have the same sign: while they agree for some families of distributions, they differ in some of the cases, and conflating them is misleading.

Many textbooks teach a rule of thumb stating that the mean is right of the median under right skew, and left of the median under left skew. This rule fails with surprising frequency. It can fail in [multimodal distributions](https://en.wikipedia.org/wiki/Multimodal_distribution), or in distributions where one tail is [long](https://en.wikipedia.org/wiki/Long_tail) but the other is [heavy](https://en.wikipedia.org/wiki/Heavy-tailed_distribution). Most commonly, though, the rule fails in discrete distributions where the areas to the left and right of the median are not equal. Such distributions not only contradict the textbook relationship between mean, median, and skew, they also contradict the textbook interpretation of the median.

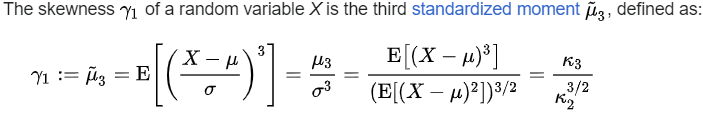
[](https://en.wikipedia.org/wiki/File:Positive_skewness_with_mean_less_than_median.png)

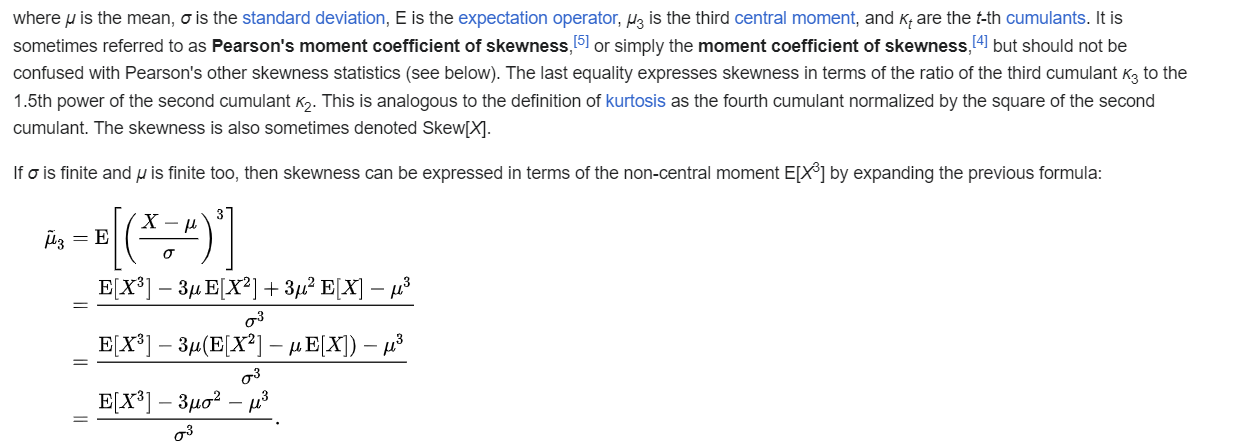
Distribution of adult residents across US households

For example, in the distribution of adult residents across US households, the skew is to the right. However, since the majority of cases is less than or equal to the mode, which is also the median, the mean sits in the heavier left tail. As a result, the rule of thumb that the mean is right of the median under right skew failed.

**Definition**

Fisher's moment coefficient of skewness

****



* **Other measures of skewness**

Other measures of skewness have been used, including simpler calculations suggested by [Karl Pearson](https://en.wikipedia.org/wiki/Karl_Pearson)

 (not to be confused with Pearson's moment coefficient of skewness, see above). These other measures are:

* Pearson's first skewness coefficient (mode skewness)

The Pearson mode skewness,[[11]](https://en.wikipedia.org/wiki/Skewness#cite_note-11) or first skewness coefficient, is defined as

[mean](https://en.wikipedia.org/wiki/Mean) − [mode](https://en.wikipedia.org/wiki/Mode_(statistics))/[standard deviation](https://en.wikipedia.org/wiki/Standard_deviation).

* Pearson's second skewness coefficient (median skewness)

The Pearson median skewness, or second skewness coefficient,[[12]](https://en.wikipedia.org/wiki/Skewness#cite_note-MWPSK-12)[[13]](https://en.wikipedia.org/wiki/Skewness#cite_note-amstat2011-13) is defined as

3 ([mean](https://en.wikipedia.org/wiki/Mean) − [median](https://en.wikipedia.org/wiki/Median))/[standard deviation](https://en.wikipedia.org/wiki/Standard_deviation).

Which is a simple multiple of the [nonparametric skew](https://en.wikipedia.org/wiki/Nonparametric_skew).