

# Average

This article has been translated from WikiSkriptä; the **translation** needs to be checked.

This article has been translated from WikiSkriptä; the **formatting** needs to be checked.

This article has been translated from WikiSkriptä; ready for the **editor's review**.

**Averages** (central tendencies) express where (around which number the data is located; <sup>[1]</sup> characterize the "center" of the data set around which the values fluctuate. The location characteristics of the sample set often estimate the true mean value of the described random variable. These include **arithmetic mean**, **geometric mean**, **median** and **mode**. Other measures of position that relate to the description of values other than the mean are **quantiles** (quartiles, deciles, percentiles, ...).

They can indicate values that the random variable itself does not even acquire. (E.g. "On average, 13.48 patients per year become ill with some disease.") Nevertheless, they have an important informative value about what the data probably looks like (or how the random variable behaves).

## Arithmetic mean

**Arithmetic mean** or **arithmetic average** ( $\bar{x}$ ) is the best-known estimate of the mean value, it is calculated as the sum of all values divided by their number:

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

The advantages of the arithmetic average are primarily easy calculation and illustrative meaning. In the case that the set has a normal distribution, or in the case of a large number of samples (measurements) from a set whose distribution has relatively mild requirements (the central limit theorem states that the set of results of random samples from such a set will have a distribution that approaching a normal distribution), the sample arithmetic mean is a consistent unbiased estimate of the mean. The main disadvantage is the considerable sensitivity of the result to outliers, e.g. if after some therapy twenty patients survive for a month and one patient survives for thirty years, the average survival is about a year and a half and the therapy can only appear to be successful based on this result. Another disadvantage is that for asymmetrically distributed data, the arithmetic mean does not make much sense.

## Weighted average

**Weighted average** or **weighted mean** is a generalization of the arithmetic mean for a statistical set in which individual values have different importance. If each element  $x_i$  is assigned a weight  $w_i$  that characterizes the significance precision of measurement, number of course credits, ...), the weighted average is defined as:

$$\bar{x} = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n x_i w_i.$$

## Geometric mean

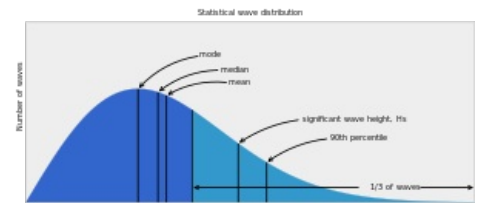
**Geometric mean** or **geometric average** ( $x_G$ ) is a more suitable alternative to the arithmetic mean when we want to express the values of a proportional variable - i.e. one that is the result of a product rather than the sum of many small effects. It only makes sense if the quantity takes mostly positive values. It is a better estimate of the mean value for data with a so-called log-normal distribution

The geometric mean of a series of  $n$  positive values of  $x_i$  is defined as the  $n$ -th square root of the product of all values.

$$x_G = \sqrt[n]{\prod_{i=1}^n x_i}.$$

If the resulting values are so low that they would disappear due to rounding, it is advisable to logarithmize the original values:

$$\ln \left( \prod_{i=1}^n x_i \right)^{1/n} = \frac{1}{n} \sum_{i=1}^n \ln x_i.$$



Examples of different ways of measuring average

Geometric average is then calculated as:

$$x_G = \exp\left(\frac{1}{n} \sum_{i=1}^n \ln x_i\right).$$

## Median

**Median** ( $\hat{x}$ ) represents the middle value of the statistical set, which is ordered from the smallest value to the largest. In the case of an even number of values, the median is the arithmetic mean of the values in the two middle places. *(In the set of numbers 1, 4, 2, 8, 11, which we sort as 1, 2, 4, 8, 11, the median will be equal to 4. In the set 1, 2, 4, 8, 11, 371, it will be the average of 4 and 8, so 6.)* The median of a random sample is a consistent (but not unbiased) estimate of the true median of a random variable. It is not as sensitive to outliers as the arithmetic mean.

## Mode

**Mode** ( $\tilde{x}$ ) represents the most frequently occurring value of the variable. *(In the set of numbers 2, 3, 5, 1, 5, 3, 7, 5, the mode number will be 5.)* It is especially important when characterizing the location of sets of qualitative data (e.g. blood groups).

## Quantiles

**Quantiles** divide a statistical set into defined parts. They are a natural generalization of the median.

- **$\alpha$ -quantile  $x_\alpha$**  (also called 100 $\alpha$ -th percentile) is a number that separates  $\alpha$  (100 $\alpha$ %) of the smallest values of a character. *(E.g. in the set of numbers 3, 4, 6, 7, 9, the twenty-five percent quantile is number 4, because 25% of the numbers have a lower value.)*
- **Percentile  $x_{0,01}$**  is a value under which is 1% of the values (1. percentile). Percentiles thus divide the set into 100 parts. *(Using the previous example, only the number 3 lies below the twenty-fifth percentile, which is the number 4. The 25%-quantile is therefore the 25th percentile.)*

Some quantiles have special nomenclature:

- **Deciles** divide the set into 10 parts. ( $x_{0,1}$  – first decile - 10-th percentile).
- **Quartiles** divide the set into 4 parts. ( $x_{0,25}$  – first quartile – 25-th percentile).
- **Median** is this *second quartile*, or *50-th percentile*.

## Links

### Related articles

- Variability rate
- Normal distribution

### External links

- Výběr a popisné statistiky, Euromise (<http://new.euromise.org/czech/tajne/ucebnice/html/html/node7.html>)
- Interaktivní sbírka úloh ze statistiky, Karel Hrach, FSE UJEP Ústí nad Labem, 2003 ([http://svareci.yc.cz/stat\\_prik\\_lady/teorie/Popisne%e1%20statistika/pr%9m%ecry%20a%20momenty.rtf](http://svareci.yc.cz/stat_prik_lady/teorie/Popisne%e1%20statistika/pr%9m%ecry%20a%20momenty.rtf))

### Literature

- 
- 

### References

- 1.

