Standard deviation
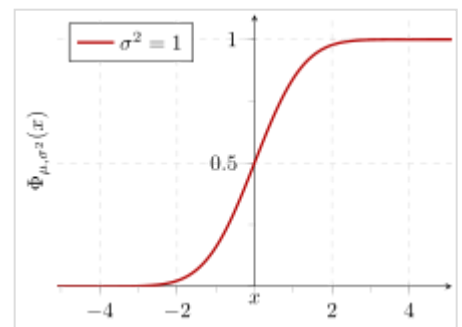
In statistics, the **standard deviation** (**SD**, also represented by the Greek letter sigma **σ** or the Latin letter s) is a measure that is used to quantify the amount of variation or dispersion of a set of data values.[1] A low standard deviation indicates that the data points tend to be close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the data points are spread out over a wider range of values.

A plot of normal distribution (or bell-shaped curve) where each band has a width of 1 standard deviation – See also: 68–95–99.7 rule

The standard deviation of a random variable, statistical population, data set, or probability distribution is the square root of its variance. It is algebraically simpler, though in practice less robust, than the average absolute deviation.[2][3] A useful property of the standard deviation is that, unlike the variance, it is expressed in the same units as the data. There are also other measures of deviation from the norm, including average absolute deviation, which provide different mathematical properties from standard deviation.[4]

Cumulative probability of a normal distribution with expected value 0 and standard deviation 1.

In addition to expressing the variability of a population, the standard deviation is commonly used to measure confidence in statistical conclusions. For example, the margin of error in polling data is determined by calculating the expected standard deviation in the results if the same poll were to be conducted multiple times. This derivation of a standard deviation is often called the "standard error" of the estimate or "standard error of the mean" when referring to a mean. It is computed as the standard deviation of all the means that would be computed from that population if an infinite number of samples were drawn and a mean for each sample were computed. It is very important to note that the standard deviation of a population and the standard error of a statistic derived from that population (such as the mean) are quite different but related (related by the inverse of the square root of the number of observations). The reported margin of error of a poll is computed from the standard error of the mean (or alternatively from the product of the standard deviation of the population and the inverse of the square root of the sample size, which is the same thing) and is typically about twice the standard deviation—the half-width of a 95 percent confidence interval. In science, many researchers report the standard deviation of experimental data, and only effects that fall much farther than two standard deviations away from what would have been expected are considered statistically

significant—normal random error or variation in the measurements is in this way distinguished from likely genuine effects or associations. The standard deviation is also important in finance, where the standard deviation on the rate of return on an investment is a measure of the volatility of the investment.

When only a sample of data from a population is available, the term **standard deviation of the sample** or **sample standard deviation** can refer to either the above-mentioned quantity as applied to those data or to a modified quantity that is an unbiased estimate of the **population standard deviation** (the standard deviation of the entire population).

Basic examples

Sample standard deviation of metabolic rate of Northern Fulmars

Logan [5] gives the following example. Furness and Bryant [6] measured the resting metabolic rate for 8 male and 6 female breeding Northern fulmars. The table shows the Furness data set.

Sex	Metabolic rate	Sex	Metabolic rate
Male	525.8	Female	727.7
Male	605.7	Female	1086.5
Male	843.3	Female	1091.0
Male	1195.5	Female	1361.3
Male	1945.6	Female	1490.5
Male	2135.6	Female	1956.1
Male	2308.7		
Male	2950.0		

The graph shows the metabolic rate for males and females. By visual inspection, it appears that the variability of the metabolic rate is greater for males than for females.

Metabolic rate versus sex for 14 northern fulmars
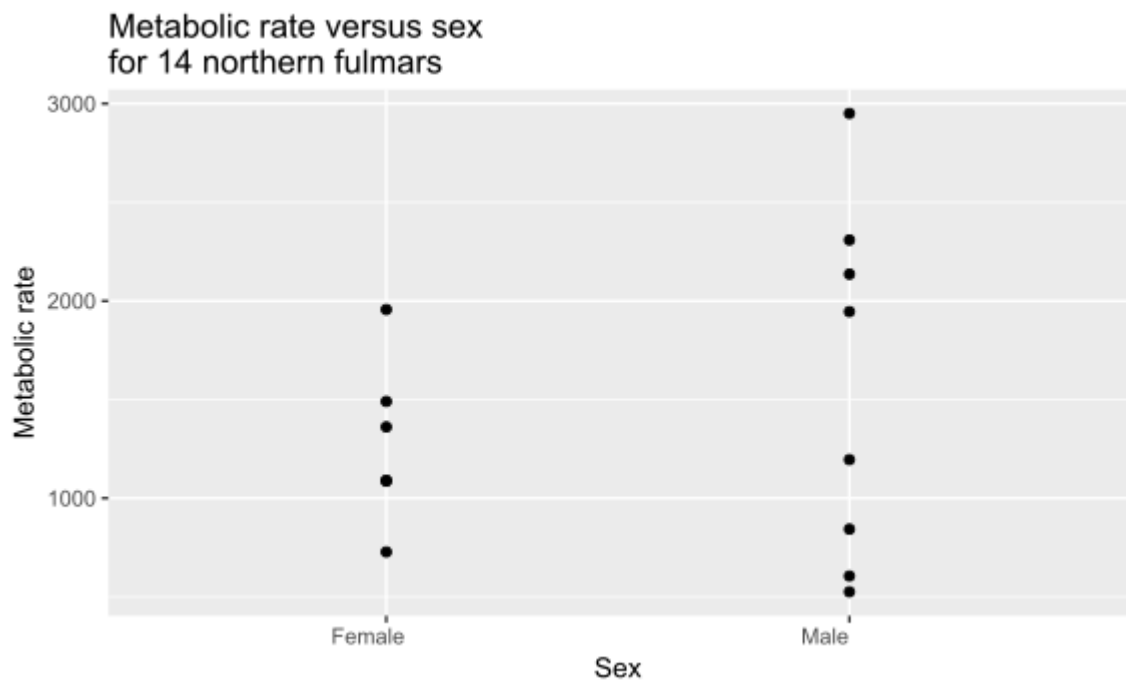
The sample standard deviation of the metabolic rate for the female fulmars is calculated as follows. The formula for the sample standard deviation is

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

where $\{x_1, x_2, \ldots, x_N\}$ are the observed values of the sample items, $\bar{x}$ is the mean value of these observations, and N is the number of observations in the sample.
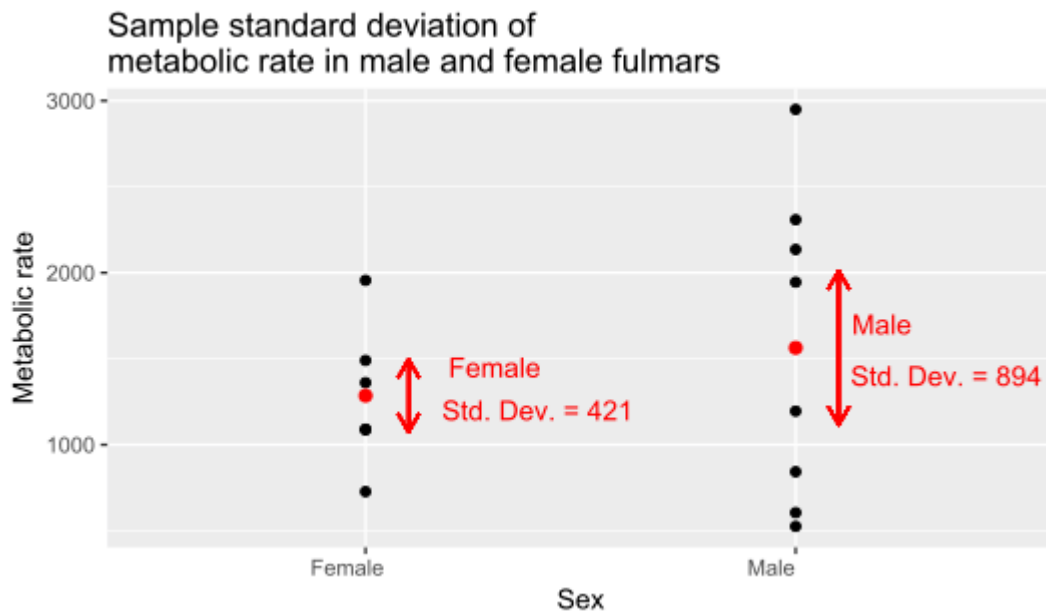
In the sample standard deviation formula, for this example, the numerator is the sum of the squared deviation of each individual animal's metabolic rate from the mean metabolic rate. The table below shows the calculation of this sum of squared deviations for the female fulmars. For females, the sum of squared deviations is 886047.09, as shown in the table.

Animal	Sex	Metabolic rate	Mean	Difference from mean	Squared difference from mean
1	Female	727.7	1285.5	-557.8	311140.84
2	Female	1086.5	1285.5	-199	39601
3	Female	1091.0	1285.5	-194.5	37830.25
4	Female	1361.3	1285.5	75.8	5745.64
5	Female	1490.5	1285.5	205	42025
6	Female	1956.1	1285.5	670.6	449704.36
	Mean =		1285.5	Sum of squared differences =	886047.09

The denominator in the sample standard deviation formula is N − 1, where N is the number of animals. In this example, there are N = 6 females, so the denominator is 6 − 1 = 5. The sample standard deviation for the female fulmars is therefore

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}} = \sqrt{\frac{886047.09}{5}} = 420.96.$$

For the male fulmars, a similar calculation gives a sample standard deviation of 894.37, approximately twice as large as the standard deviation for the females. The graph shows the metabolic rate data, the means (red dots), and the standard deviations (red lines) for females and males.

Sample standard deviation of metabolic rate in male and female fulmars

Use of the sample standard deviation implies that these 14 fulmars are a sample from a larger population of fulmars. If these 14 fulmars comprised the entire population (perhaps the last 14 surviving fulmars), then instead of the sample standard deviation, the calculation would use the population standard deviation. In the population standard deviation formula, the denominator is N instead of N-1. It is rare that measurements can be taken for an entire population, so, by default, statistical software packages calculate the sample standard deviation. Similarly, journal articles report the sample standard deviation unless otherwise specified.

Population standard deviation of grades of eight students

Suppose that the entire population of interest was eight students in a particular class. For a finite set of numbers, the population standard deviation is found by taking the square root of the average of the squared deviations of the values from their average value. The marks of a class of eight students (that is, a statistical population) are the following eight values:

2, 4, 4, 4, 5, 5, 7, 9.

These eight data points have the mean (average) of 5:

$$\frac{2 + 4 + 4 + 4 + 5 + 5 + 7 + 9}{8} = 5.$$

First, calculate the deviations of each data point from the mean, and square the result of each:

$$(2 - 5)^2 = (-3)^2 = 9 \qquad (5 - 5)^2 = 0^2 = 0$$
$$(4 - 5)^2 = (-1)^2 = 1 \qquad (5 - 5)^2 = 0^2 = 0$$
$$(4 - 5)^2 = (-1)^2 = 1 \qquad (7 - 5)^2 = 2^2 = 4$$
$$(4 - 5)^2 = (-1)^2 = 1 \qquad (9 - 5)^2 = 4^2 = 16.$$

The variance is the mean of these values:

$$\frac{9 + 1 + 1 + 1 + 0 + 0 + 4 + 16}{8} = 4.$$

and the *population* standard deviation is equal to the square root of the variance:

$$\sqrt{4} = 2.$$

This formula is valid only if the eight values with which we began form the complete population. If the values instead were a random sample drawn from some large parent population (for example, they were 8 marks randomly and independently chosen from a class of 2 million), then one often divides by 7 (which is $n - 1$) instead of 8 (which is n) in the denominator of the last formula. In that case the result of the original formula would be called the *sample* standard deviation. Dividing by $n - 1$ rather than by n gives an unbiased estimate of the variance of the larger parent population. This is known as *Bessel's correction*.[7]

Standard deviation of average height for adult men

If the population of interest is approximately normally distributed, the standard deviation provides information on the proportion of observations above or below certain values. For example, the average height for adult men in the United States is about 70 inches (177.8 cm), with a standard deviation of around 3 inches (7.62 cm). This means that most men (about 68%, assuming a normal distribution) have a height within 3 inches (7.62 cm) of the mean (67–73 inches (170.18–185.42 cm)) – one standard deviation – and almost all men (about 95%) have a height within 6 inches (15.24 cm) of the mean (64–76 inches (162.56–193.04 cm)) – two standard deviations. If the standard deviation were zero, then all men would be exactly 70 inches (177.8 cm) tall. If the standard deviation were 20 inches (50.8 cm), then men would have much more variable heights, with a typical range of about 50–90 inches (127–228.6 cm). Three standard deviations account for 99.7% of the sample population being studied, assuming the distribution is normal (bell-shaped). (See the 68-95-99.7 rule, or the *empirical rule,* for more information.)

Definition of population values

Let X be a random variable with mean value μ:

$$\mathrm{E}[X] = \mu.$$

Here the operator E denotes the average or expected value of X. Then the **standard deviation** of X is the quantity

$$
\begin{aligned}
\sigma &= \sqrt{\mathrm{E}[(X - \mu)^2]} \\
&= \sqrt{\mathrm{E}[X^2] + \mathrm{E}[-2\mu X] + \mathrm{E}[\mu^2]} \\
&= \sqrt{\mathrm{E}[X^2] - 2\mu \, \mathrm{E}[X] + \mu^2} \\
&= \sqrt{\mathrm{E}[X^2] - 2\mu^2 + \mu^2} \\
&= \sqrt{\mathrm{E}[X^2] - \mu^2} \\
&= \sqrt{\mathrm{E}[X^2] - (\mathrm{E}[X])^2}
\end{aligned}
$$

(derived using the properties of expected value).

In other words, the standard deviation σ (sigma) is the square root of the variance of X; i.e., it is the square root of the average value of $(X - \mu)^2$.

The standard deviation of a (univariate) probability distribution is the same as that of a random variable having that distribution. Not all random variables have a standard deviation, since these expected values need not exist. For example, the standard deviation of a random variable that follows a Cauchy distribution is undefined because its expected value μ is undefined.

Discrete random variable

In the case where X takes random values from a finite data set $x_1, x_2, ..., x_N$, with each value having the same probability, the standard deviation is

$$\sigma = \sqrt{\frac{1}{N}\left[(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_N - \mu)^2\right]}, \quad \text{where} \ \mu = \frac{1}{N}(x_1 + \cdots + x_N),$$

or, using summation notation,

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}, \quad \text{where} \ \mu = \frac{1}{N}\sum_{i=1}^{N}x_i.$$

If, instead of having equal probabilities, the values have different probabilities, let x_1 have probability p_1, x_2 have probability p_2, ..., x_N have probability p_N. In this case, the standard deviation will be

$$\sigma = \sqrt{\sum_{i=1}^{N}p_i(x_i - \mu)^2}, \quad \text{where} \ \mu = \sum_{i=1}^{N}p_i x_i.$$

Continuous random variable

The standard deviation of a continuous real-valued random variable X with probability density function $p(x)$ is

$$\sigma = \sqrt{\int_{\mathbf{X}}(x - \mu)^2\, p(x)\, \mathrm{d}x}, \quad \text{where} \ \mu = \int_{\mathbf{X}} x\, p(x)\, \mathrm{d}x,$$

and where the integrals are definite integrals taken for x ranging over the set of possible values of the random variable X.

In the case of a parametric family of distributions, the standard deviation can be expressed in terms of the parameters. For example, in the case of the log-normal distribution with parameters μ and σ^2, the standard deviation is $[(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)]^{1/2}$.

Estimation

One can find the standard deviation of an entire population in cases (such as standardized testing) where every member of a population is sampled. In cases where that cannot be done, the standard deviation σ is estimated by examining a random sample taken from the population and computing a statistic of the sample, which is used as an estimate of the population standard deviation. Such a statistic is called an estimator, and the estimator (or the value of the estimator, namely the estimate) is called a **sample standard deviation,** and is denoted by s (possibly with modifiers). However, unlike in the case of estimating the population mean, for which the sample mean is a simple estimator with many desirable properties (unbiased, efficient, maximum likelihood), there is no single estimator for the standard deviation with all these properties, and unbiased estimation of standard deviation is a very technically involved problem. Most often, the standard deviation is estimated using the *corrected sample standard deviation* (using $N - 1$), defined below, and this is often referred to as the "sample standard deviation", without qualifiers. However, other estimators are better in other respects: the uncorrected estimator (using N) yields lower mean squared error, while using $N - 1.5$ (for the normal distribution) almost completely eliminates bias.

Uncorrected sample standard deviation

The formula for the *population* standard deviation (of a finite population) can be applied to the sample, using the size of the sample as the size of the population (though the actual population size from which the sample is drawn may be much larger). This estimator, denoted by s_N, is known as the **uncorrected sample standard deviation**, or sometimes the **standard deviation of the sample** (considered as the entire population), and is defined as follows:

$$s_N = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2},$$

where $\{x_1, x_2, \ldots, x_N\}$ are the observed values of the sample items and $\overline{x}$ is the mean value of these observations, while the denominator N stands for the size of the sample: this is the square root of the sample variance, which is the average of the squared deviations about the sample mean.

This is a consistent estimator (it converges in probability to the population value as the number of samples goes to infinity), and is the maximum-likelihood estimate when the population is normally distributed. However, this is a biased estimator, as the estimates are generally too low. The bias decreases as sample size grows, dropping off as $1/N$, and thus is most significant for small or moderate sample sizes; for $N > 75$ the bias is below 1%. Thus for very large sample sizes, the uncorrected sample standard deviation is generally acceptable. This estimator also has a uniformly smaller mean squared error than the corrected sample standard deviation.

Corrected sample standard deviation

If the *biased sample variance* (the second central moment of the sample, which is a downward-biased estimate of the population variance) is used to compute an estimate of the population's standard deviation, the result is

$$s_N = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2}.$$

Here taking the square root introduces further downward bias, by Jensen's inequality, due to the square root being a concave function. The bias in the variance is easily corrected, but the bias from the square root is more difficult to correct, and depends on the distribution in question.

An unbiased estimator for the *variance* is given by applying Bessel's correction, using $N - 1$ instead of N to yield the *unbiased sample variance,* denoted s^2:

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

This estimator is unbiased if the variance exists and the sample values are drawn independently with replacement. $N - 1$ corresponds to the number of degrees of freedom in the vector of deviations from the mean, $(x_1 - \bar{x}, \ldots, x_n - \bar{x})$.

Taking square roots reintroduces bias (because the square root is a nonlinear function, which does not commute with the expectation), yielding the **corrected sample standard deviation,** denoted by s:

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

As explained above, while s^2 is an unbiased estimator for the population variance, s is still a biased estimator for the population standard deviation, though markedly less biased than the uncorrected sample standard deviation. This estimator is commonly used and generally known simply as the "sample standard deviation". The bias may still be large for small samples (N less than 10). As sample size increases, the amount of bias decreases. We obtain more information and the difference between $\frac{1}{N}$ and $\frac{1}{N-1}$ becomes smaller.

Unbiased sample standard deviation

For unbiased estimation of standard deviation, there is no formula that works across all distributions, unlike for mean and variance. Instead, s is used as a basis, and is scaled by a correction factor to produce an unbiased estimate. For the normal distribution, an unbiased estimator is given by s/c_4, where the correction factor (which depends on N) is given in terms of the Gamma function, and equals:

$$c_4(N) = \sqrt{\frac{2}{N-1}} \, \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)}.$$

This arises because the sampling distribution of the sample standard deviation follows a (scaled) chi distribution, and the correction factor is the mean of the chi distribution.

An approximation can be given by replacing $N - 1$ with $N - 1.5$, yielding:

$$\hat{\sigma} = \sqrt{\frac{1}{N-1.5} \sum_{i=1}^{N} (x_i - \bar{x})^2},$$

The error in this approximation decays quadratically (as $1/N^2$), and it is suited for all but the smallest samples or highest precision: for n = 3 the bias is equal to 1.3%, and for n = 9 the bias is already less than 0.1%.

For other distributions, the correct formula depends on the distribution, but a rule of thumb is to use the further refinement of the approximation:

$$\hat{\sigma} = \sqrt{\frac{1}{N - 1.5 - \frac{1}{4}\gamma_2} \sum_{i=1}^{N}(x_i - \bar{x})^2},$$

where γ_2 denotes the population excess kurtosis. The excess kurtosis may be either known beforehand for certain distributions, or estimated from the data.

Confidence interval of a sampled standard deviation

The standard deviation we obtain by sampling a distribution is itself not absolutely accurate, both for mathematical reasons (explained here by the confidence interval) and for practical reasons of measurement (measurement error). The mathematical effect can be described by the confidence interval or CI. To show how a larger sample will make the confidence interval narrower, consider the following examples: A small population of N = 2 has only 1 degree of freedom for estimating the standard deviation. The result is that a 95% CI of the SD runs from 0.45 × SD to 31.9 × SD; the factors here are as follows:

$$\Pr\left(q_{\alpha/2} < k\frac{s^2}{\sigma^2} < q_{1-\alpha/2}\right) = 1 - \alpha,$$

where q_p is the p-th quantile of the chi-square distribution with k degrees of freedom, and $1 - \alpha$ is the confidence level. This is equivalent to the following:

$$\Pr\left(k\frac{s^2}{q_{1-\alpha/2}} < \sigma^2 < k\frac{s^2}{q_{\alpha/2}}\right) = 1 - \alpha.$$

With k = 1, $q_{0.025} = 0.000982$ and $q_{0.975} = 5.024$. The reciprocals of the square roots of these two numbers give us the factors 0.45 and 31.9 given above.

A larger population of N = 10 has 9 degrees of freedom for estimating the standard deviation. The same computations as above give us in this case a 95% CI running from 0.69*SD to 1.83*SD. So even with a sample population of 10, the actual SD can still be almost a factor 2 higher than the sampled SD. For a sample population N=100, this is down to 0.88*SD to 1.16*SD. To be more certain that the sampled SD is close to the actual SD we need to sample a large number of points.

These same formulae can be used to obtain confidence intervals on the variance of residuals from a least squares fit under standard normal theory, where k is now the number of degrees of freedom for error.

Identities and mathematical properties

The standard deviation is invariant under changes in location, and scales directly with the scale of the random variable. Thus, for a constant c and random variables X and Y:

$$\sigma(c) = 0$$

$$\sigma(X + c) = \sigma(X),$$

$$\sigma(cX) = |c|\sigma(X).$$

The standard deviation of the sum of two random variables can be related to their individual standard deviations and the covariance between them:

$$\sigma(X + Y) = \sqrt{\mathrm{var}(X) + \mathrm{var}(Y) + 2\ \mathrm{cov}(X, Y)}.$$

where $\mathrm{var} = \sigma^2$ and cov stand for variance and covariance, respectively.

The calculation of the sum of squared deviations can be related to moments calculated directly from the data. In the following formula, the letter E is interpreted to mean expected value, i.e., mean.

$$\sigma(X) = \sqrt{E[(X - E(X))^2]} = \sqrt{E[X^2] - (E[X])^2}.$$

The sample standard deviation can be computed as:

$$s(X) = \sqrt{\frac{N}{N-1}}\sqrt{E[(X - E(X))^2]}.$$

For a finite population with equal probabilities at all points, we have

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

This means that the standard deviation is equal to the square root of the difference between the average of the squares of the values and the square of the average value. See computational formula for the variance for proof, and for an analogous result for the sample standard deviation.

Interpretation and application

A large standard deviation indicates that the data points can spread far from the mean and a small standard deviation indicates that they are clustered closely around the mean.

For example, each of the three populations {0, 0, 14, 14}, {0, 6, 8, 14} and {6, 6, 8, 8} has a mean of 7. Their standard deviations are 7, 5, and 1, respectively. The third population has a much smaller standard deviation than the other two because its values are all close to 7. It will have the same units as the data points themselves. If, for instance, the data set {0, 6, 8, 14} represents the ages of a population of four siblings in years, the standard deviation is 5 years. As another example, the population {1000, 1006, 1008, 1014} may represent the distances traveled by four athletes, measured in meters. It has a mean of 1007 meters, and a standard deviation of 5 meters.

Standard deviation may serve as a measure of uncertainty. In physical science, for example, the reported standard deviation of a group of repeated underlined measurements gives the underlined precision of those measurements. When deciding whether measurements agree with a theoretical prediction, the standard deviation of those measurements is of crucial importance: if the mean of the measurements is too far away from the prediction (with the distance measured in standard deviations), then the theory being tested probably needs to be revised. This makes sense since they fall outside the range of values that could reasonably be expected to occur, if the prediction were correct and the standard deviation appropriately quantified. See prediction interval.
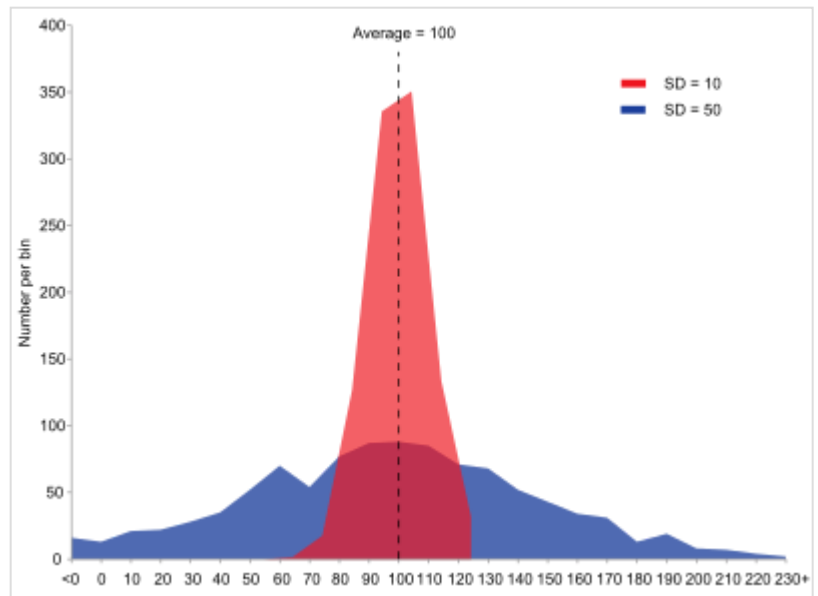
Example of samples from two populations with the same mean but different standard deviations. Red population has mean 100 and SD 10; blue population has mean 100 and SD 50.

While the standard deviation does measure how far typical values tend to be from the mean, other measures are available. An example is the mean absolute deviation, which might be considered a more direct measure of average distance, compared to the root mean square distance inherent in the standard deviation.

Application examples

The practical value of understanding the standard deviation of a set of values is in appreciating how much variation there is from the average (mean).

Experiment, industrial and hypothesis testing

Standard deviation is often used to compare real-world data against a model to test the model. For example, in industrial applications the weight of products coming off a production line may need to comply with a legally required value. By weighing some fraction of the products an average weight can be found, which will always be slightly different to the long-term average. By using standard deviations, a minimum and maximum value can be calculated that the averaged weight will be within some very high percentage of the time (99.9% or more). If it falls outside the range then the production process may need to be corrected. Statistical tests such as these are particularly important when the testing is relatively expensive. For example, if the product needs to be opened and drained and weighed, or if the product was otherwise used up by the test.

In experimental science, a theoretical model of reality is used. Particle physics conventionally uses a standard of "5 sigma" for the declaration of a discovery.[8] A five-sigma level translates to one chance in 3.5 million that a random fluctuation would yield the result. This level of certainty was required in order to assert that a particle consistent with the Higgs boson had been discovered in two independent experiments at CERN,[9] and this was also the significance level leading to the declaration of the first detection of gravitational waves.[10]

Weather

As a simple example, consider the average daily maximum temperatures for two cities, one inland and one on the coast. It is helpful to understand that the range of daily maximum temperatures for cities near the coast is smaller than for cities inland. Thus, while these two cities may each have the same average maximum temperature, the standard deviation of the daily maximum temperature for the coastal city will be less than that of the inland city as, on any particular day, the actual maximum temperature is more likely to be farther from the average maximum temperature for the inland city than for the coastal one.

Finance

In finance, standard deviation is often used as a measure of the risk associated with price-fluctuations of a given asset (stocks, bonds, property, etc.), or the risk of a portfolio of assets[11] (actively managed mutual funds, index mutual funds, or ETFs). Risk is an important factor in determining how to efficiently manage a portfolio of investments because it determines the variation in returns on the asset and/or portfolio and gives investors a mathematical basis for investment decisions (known as mean-variance optimization). The fundamental concept of risk is that as it increases, the expected return on an investment should increase as well, an increase known as the risk premium. In other words, investors should expect a higher return on an investment when that investment carries a higher level of risk or uncertainty. When evaluating investments, investors should estimate both the expected return and the uncertainty of future returns. Standard deviation provides a quantified estimate of the uncertainty of future returns.

For example, assume an investor had to choose between two stocks. Stock A over the past 20 years had an average return of 10 percent, with a standard deviation of 20 percentage points (pp) and Stock B, over the same period, had average returns of 12 percent but a higher standard deviation of 30 pp. On the basis of risk and return, an investor may decide that Stock A is the safer choice, because Stock B's additional two percentage points of return is not worth the additional 10 pp standard deviation (greater risk or uncertainty of the expected return). Stock B is likely to fall short of the initial investment (but also to exceed the initial investment) more often than Stock A under the same circumstances, and is estimated to return only two percent more on average. In this example, Stock A is expected to earn about 10 percent, plus or minus 20 pp (a range of 30 percent to −10 percent), about two-thirds of the future year returns. When considering more extreme possible returns or outcomes in future, an investor should expect results of as much as 10 percent plus or minus 60 pp, or a range from 70 percent to −50 percent, which includes outcomes for three standard deviations from the average return (about 99.7 percent of probable returns).

Calculating the average (or arithmetic mean) of the return of a security over a given period will generate the expected return of the asset. For each period, subtracting the expected return from the actual return results in the difference from the mean. Squaring the difference in each period and taking the average gives the overall variance of the return of the asset. The larger the variance, the greater risk the security carries. Finding the square root of this variance will give the standard deviation of the investment tool in question.

Population standard deviation is used to set the width of Bollinger Bands, a widely adopted technical analysis tool. For example, the upper Bollinger Band is given as $\bar{x} + n\sigma_x$. The most commonly used value for n is 2; there is about a five percent chance of going outside, assuming a normal distribution of returns.

Financial time series are known to be non-stationary series, whereas the statistical calculations above, such as standard deviation, apply only to stationary series. To apply the above statistical tools to non-stationary series, the series first must be transformed to a stationary series, enabling use of statistical tools that now have a valid basis from which to work.

Geometric interpretation

To gain some geometric insights and clarification, we will start with a population of three values, x_1, x_2, x_3. This defines a point $P = (x_1, x_2, x_3)$ in $\mathbf{R}^3$. Consider the line $L = \{(r, r, r) : r \in \mathbf{R}\}$. This is the "main diagonal" going through the origin. If our three given values were all equal, then the standard deviation would be zero and P would lie on L. So it is not unreasonable to assume that the standard deviation is related to the *distance* of P to L. That is indeed the case. To move orthogonally from L to the point P, one begins at the point:

$$M = (\overline{x}, \overline{x}, \overline{x})$$

whose coordinates are the mean of the values we started out with.

Derivation of $M = (\overline{x}, \overline{x}, \overline{x})$	[show]

M is on L therefore $M = (l, l, l)$ with $l \in \mathbf{R}$

The line L is to be orthogonal to the vector from M to P. Therefore:

$$L \cdot (P - M) = 0$$
$$(r, r, r) \cdot (x_1 - l, x_2 - l, x_3 - l) = 0$$
$$r * (x_1 - l + x_2 - l + x_3 - l) = 0$$
$$r * \left(\sum_i x_i - 3l\right) = 0$$
$$\sum_i x_i - 3l = 0$$
$$\frac{1}{3}\sum_i x_i = l$$
$$\overline{x} = l$$

A little algebra shows that the distance between P and M (which is the same as the orthogonal distance between P and the line L) $\sqrt{\sum_i (x_i - \overline{x})^2}$ is equal to the standard deviation of the vector x_1, x_2, x_3, multiplied by the square root of the number of dimensions of the vector (3 in this case.)

Chebyshev's inequality

An observation is rarely more than a few standard deviations away from the mean. Chebyshev's inequality ensures that, for all distributions for which the standard deviation is defined, the amount of data within a number of standard deviations of the mean is at least as much as given in the following table.

Distance from mean	Minimum population
$\sqrt{2}\,\sigma$	50%
2σ	75%
3σ	89%
4σ	94%
5σ	96%
6σ	97%
$k\sigma$	$1 - \dfrac{1}{k^2}$ [12]
$\dfrac{1}{\sqrt{1-\ell}}\,\sigma$	ℓ

Rules for normally distributed data

The central limit theorem states that the distribution of an average of many independent, identically distributed random variables tends toward the famous bell-shaped normal distribution with a probability density function of

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where μ is the expected value of the random variables, σ equals their distribution's standard deviation divided by $n^{1/2}$, and n is the number of random variables. The standard deviation therefore is simply a scaling variable that adjusts how broad the curve will be, though it also appears in the normalizing constant.
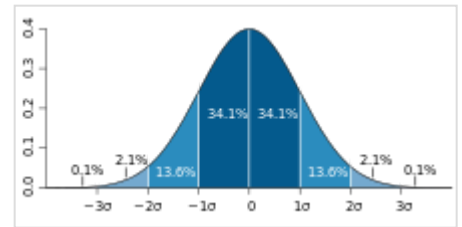
If a data distribution is approximately normal, then the proportion of data values within z standard deviations of the mean is defined by:

$$\text{Proportion} = \text{erf}\left(\frac{z}{\sqrt{2}}\right)$$

Dark blue is one standard deviation on either side of the mean. For the normal distribution, this accounts for 68.27 percent of the set; while two standard deviations from the mean (medium and dark blue) account for 95.45 percent; three standard deviations (light, medium, and dark blue) account for 99.73 percent; and four standard deviations account for 99.994 percent. The two points of the curve that are one standard deviation from the mean are also the inflection points.

where erf is the error function. The proportion that is less than or equal to a number, x, is given by the cumulative distribution function:

$$\text{Proportion} \leq x = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] = \frac{1}{2}\left[1 + \text{erf}\left(\frac{z}{\sqrt{2}}\right)\right].\text{[13]}$$

If a data distribution is approximately normal then about 68 percent of the data values are within one standard deviation of the mean (mathematically, $\mu \pm \sigma$, where μ is the arithmetic mean), about 95 percent are within two standard deviations ($\mu \pm 2\sigma$), and about 99.7 percent lie within three standard deviations ($\mu \pm 3\sigma$). This is known as the *68-95-99.7 rule*, or *the empirical rule*.

For various values of z, the percentage of values expected to lie in and outside the symmetric interval, CI = $(-z\sigma, z\sigma)$, are as follows:
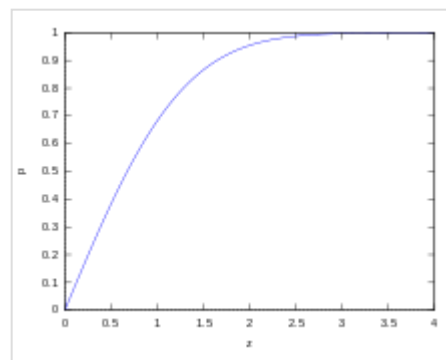
Percentage within(z)
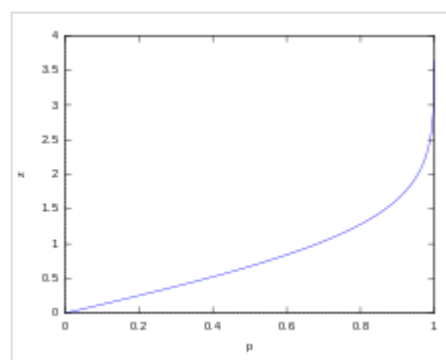
z(Percentage within)

Confidence interval	Proportion within	Proportion without	
	Percentage	Percentage	Fraction
0.674 490σ	50%	50%	1 / 2
0.994 458σ	68%	32%	1 / 3.125
1σ	68.268 9492%	31.731 0508%	1 / 3.151 4872
1.281 552σ	80%	20%	1 / 5
1.644 854σ	90%	10%	1 / 10
1.959 964σ	95%	5%	1 / 20
2σ	95.449 9736%	4.550 0264%	1 / 21.977 895
2.575 829σ	99%	1%	1 / 100
3σ	99.730 0204%	0.269 9796%	1 / 370.398
3.290 527σ	99.9%	0.1%	1 / 1000
3.890 592σ	99.99%	0.01%	1 / 10 000
4σ	99.993 666%	0.006 334%	1 / 15 787
4.417 173σ	99.999%	0.001%	1 / 100 000
4.5σ	*99.999 320 465 3751%*	*0.000 679 534 6249%*	*3.4 / 1 000 000 (on each side of mean)*
4.891 638σ	99.9999%	0.0001%	1 / 1 000 000
5σ	99.999 942 6697%	0.000 057 3303%	1 / 1 744 278
5.326 724σ	99.999 99%	0.000 01%	1 / 10 000 000
5.730 729σ	99.999 999%	0.000 001%	1 / 100 000 000
6σ	99.999 999 8027%	0.000 000 1973%	1 / 506 797 346
6.109 410σ	99.999 9999%	0.000 0001%	1 / 1 000 000 000
6.466 951σ	99.999 999 99%	0.000 000 01%	1 / 10 000 000 000
6.806 502σ	99.999 999 999%	0.000 000 001%	1 / 100 000 000 000
7σ	99.999 999 999 7440%	0.000 000 000 256%	1 / 390 682 215 445

Relationship between standard deviation and mean

The mean and the standard deviation of a set of data are descriptive statistics usually reported together. In a certain sense, the standard deviation is a "natural" measure of statistical dispersion if the center of the data is measured about the mean. This is because the standard deviation from the mean is smaller than from any other point. The precise statement is the following: suppose x_1, ..., x_n are real numbers and define the function:

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

Using calculus or by completing the square, it is possible to show that $\sigma(r)$ has a unique minimum at the mean:

$$r = \bar{x}.$$

Variability can also be measured by the <u>coefficient of variation</u>, which is the ratio of the standard deviation to the mean. It is a <u>dimensionless number</u>.

Standard deviation of the mean

Often, we want some information about the precision of the mean we obtained. We can obtain this by determining the standard deviation of the sampled mean. Assuming statistical independence of the values in the sample, the standard deviation of the mean is related to the standard deviation of the distribution by:

$$\sigma_{\text{mean}} = \frac{1}{\sqrt{N}}\sigma$$

where N is the number of observations in the sample used to estimate the mean. This can easily be proven with (see <u>basic properties of the variance</u>):

$$\text{var}(X) \equiv \sigma_X^2$$
$$\text{var}(X_1 + X_2) \equiv \text{var}(X_1) + \text{var}(X_2)$$

(Statistical Independence is assumed.)

$$\text{var}(cX_1) \equiv c^2 \text{ var}(X_1)$$

hence

$$\text{var}(\text{mean}) = \text{var}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{1}{N^2}\text{var}\left(\sum_{i=1}^{N} X_i\right)$$
$$= \frac{1}{N^2}\sum_{i=1}^{N}\text{var}(X_i) = \frac{N}{N^2}\text{var}(X) = \frac{1}{N}\text{var}(X).$$

Resulting in:

$$\sigma_{\text{mean}} = \frac{\sigma}{\sqrt{N}}.$$

It should be emphasized that in order to estimate the standard deviation of the mean σ_{mean} it is necessary to know the standard deviation of the entire population σ beforehand. However, in most applications this parameter is unknown. For example, if a series of 10 measurements of a previously unknown quantity is performed in a laboratory, it is possible to calculate the resulting sample mean and sample standard deviation, but it is impossible to calculate the standard deviation of the mean.

Rapid calculation methods

The following two formulas can represent a running (repeatedly updated) standard deviation. A set of two power sums s_1 and s_2 are computed over a set of N values of x, denoted as $x_1, ..., x_N$:

$$s_j = \sum_{k=1}^{N} x_k^j.$$

Given the results of these running summations, the values N, s_1, s_2 can be used at any time to compute the *current* value of the running standard deviation:

$$\sigma = \frac{\sqrt{Ns_2 - s_1^2}}{N}$$

Where N, as mentioned above, is the size of the set of values (or can also be regarded as s_0).

Similarly for sample standard deviation,

$$s = \sqrt{\frac{Ns_2 - s_1^2}{N(N-1)}}.$$

In a computer implementation, as the three s_j sums become large, we need to consider round-off error, arithmetic overflow, and arithmetic underflow. The method below calculates the running sums method with reduced rounding errors.[14] This is a "one pass" algorithm for calculating variance of n samples without the need to store prior data during the calculation. Applying this method to a time series will result in successive values of standard deviation corresponding to n data points as n grows larger with each new sample, rather than a constant-width sliding window calculation.

For $k = 1, ..., n$:

$$A_0 = 0$$
$$A_k = A_{k-1} + \frac{x_k - A_{k-1}}{k}$$

where A is the mean value.

$$Q_0 = 0$$
$$Q_k = Q_{k-1} + \frac{k-1}{k}(x_k - A_{k-1})^2 = Q_{k-1} + (x_k - A_{k-1})(x_k - A_k)$$

Note: $Q_1 = 0$ since $k - 1 = 0$ or $x_1 = A_1$

Sample variance:

$$s_n^2 = \frac{Q_n}{n-1}$$

Population variance:

$$\sigma_n^2 = \frac{Q_n}{n}$$

Weighted calculation

When the values x_i are weighted with unequal weights w_i, the power sums s_0, s_1, s_2 are each computed as:

$$s_j = \sum_{k=1}^{N} w_k x_k^j.$$

And the standard deviation equations remain unchanged. Note that s_0 is now the sum of the weights and not the number of samples N.

The incremental method with reduced rounding errors can also be applied, with some additional complexity.

A running sum of weights must be computed for each k from 1 to n:

$$W_0 = 0$$
$$W_k = W_{k-1} + w_k$$

and places where $1/n$ is used above must be replaced by w_i/W_n:

$$A_0 = 0$$
$$A_k = A_{k-1} + \frac{w_k}{W_k}(x_k - A_{k-1})$$
$$Q_0 = 0$$
$$Q_k = Q_{k-1} + \frac{w_k W_{k-1}}{W_k}(x_k - A_{k-1})^2 = Q_{k-1} + w_k(x_k - A_{k-1})(x_k - A_k)$$

In the final division,

$$\sigma_n^2 = \frac{Q_n}{W_n}$$

and

$$s_n^2 = \frac{Q_n}{W_n - 1},$$

or

$$s_n^2 = \frac{n'}{n'-1}\sigma_n^2,$$

where n is the total number of elements, and n' is the number of elements with non-zero weights. The above formulas become equal to the simpler formulas given above if weights are taken as equal to one.

History

The term *standard deviation* was first used[15] in writing by <u>Karl Pearson</u>[16] in 1894, following his use of it in lectures. This was as a replacement for earlier alternative names for the same idea: for example, <u>Gauss</u> used *mean error*.[17]

See also

- <u>68–95–99.7 rule</u>
- <u>Accuracy and precision</u>

- Chebyshev's inequality An inequality on location and scale parameters
- Cumulant
- Deviation (statistics)
- Distance correlation Distance standard deviation
- Error bar
- Geometric standard deviation
- Mahalanobis distance generalizing number of standard deviations to the mean
- Mean absolute error
- Pooled standard deviation
- Propagation of uncertainty
- Percentile
- Raw score
- Relative standard deviation
- Robust standard deviation
- Root mean square
- Sample size
- Samuelson's inequality
- Six Sigma
- Standard error
- Standard score
- Volatility (finance)
- Yamartino method for calculating standard deviation of wind direction

References

1. Bland, J.M.; Altman, D.G. (1996). "Statistics notes: measurement error" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2351401). *BMJ*. **312** (7047): 1654. doi:10.1136/bmj.312.7047.1654 (https://doi.org/10.1136%2Fbmj.312.7047.1654). PMC 2351401 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2351401). PMID 8664723 (https://pubmed.ncbi.nlm.nih.gov/8664723).
2. Gauss, Carl Friedrich (1816). "Bestimmung der Genauigkeit der Beobachtungen". *Zeitschrift für Astronomie und verwandte Wissenschaften*. **1**: 187–197.
3. Walker, Helen (1931). *Studies in the History of the Statistical Method*. Baltimore, MD: Williams & Wilkins Co. pp. 24–25.
4. Gorard, Stephen. Revisiting a 90-year-old debate: the advantages of the mean deviation (http://www.leeds.ac.uk/educol/documents/00003759.htm). Department of Educational Studies, University of York
5. Logan, Murray (2010), *Biostatistical Design and Analysis Using R* (http://users.monash.edu.au/~murray/BDAR/) (First ed.), Wiley-Blackwell
6. Furness, R.W.; Bryant, D.M. (1996). "Effect of wind on field metabolic rates of breeding Northern fulmars". *Ecology*. **77**: 1181–1188.
7. Weisstein, Eric W. "Bessel's Correction" (https://mathworld.wolfram.com/BesselsCorrection.html). *MathWorld*.
8. "CERN | Accelerating science" (http://public.web.cern.ch/public/). Public.web.cern.ch. Retrieved 10 August 2013.
9. "CERN experiments observe particle consistent with long-sought Higgs boson | CERN press office" (http://press-archive.web.cern.ch/press-archive/PressReleases/Releases2012/PR17.12E.html). Press.web.cern.ch. 4 July 2012. Retrieved 30 May 2015.
10. LIGO Scientific Collaboration, Virgo Collaboration (2016), "Observation of Gravitational Waves from a Binary Black Hole Merger", *Physical Review Letters*, **116** (6): 061102, arXiv:1602.03837 (https://arxiv.org/abs/1602.03837), Bibcode:2016PhRvL.116f1102A (https://ui.adsabs.harvard.edu/abs/2016PhRvL.116f1102A), doi:10.1103/PhysRevLett.116.061102 (https://doi.org/10.1103%2FPhysRevLett.116.061102), PMID 26918975 (https://pubmed.ncbi.nlm.nih.gov/26918975)
11. "What is Standard Deviation" (http://www.edupristine.com/blog/what-is-standard-deviation). Pristine. Retrieved 29 October 2011.
12. Ghahramani, Saeed (2000). *Fundamentals of Probability* (2nd Edition). Prentice Hall: New Jersey. p. 438.

13. Eric W. Weisstein. "Distribution Function" (http://mathworld.wolfram.com/DistributionFunction.html). MathWorld—A Wolfram Web Resource. Retrieved 30 September 2014.
14. Welford, BP (August 1962). "Note on a Method for Calculating Corrected Sums of Squares and Products" (http://zach.in.tu-clausthal.de/teaching/info_literatur/Welford.pdf) (PDF). *Technometrics*. **4** (3): 419–420. doi:10.1080/00401706.1962.10490022 (https://doi.org/10.1080%2F00401706.1962.10490022).
15. Dodge, Yadolah (2003). *The Oxford Dictionary of Statistical Terms*. Oxford University Press. ISBN 0-19-920613-9.
16. Pearson, Karl (1894). "On the dissection of asymmetrical frequency curves". *Philosophical Transactions of the Royal Society A*. **185**: 71–110. Bibcode:1894RSPTA.185...71P (https://ui.adsabs.harvard.edu/abs/1894RSPTA.185...71P). doi:10.1098/rsta.1894.0003 (https://doi.org/10.1098%2Frsta.1894.0003).
17. Miller, Jeff. "Earliest Known Uses of Some of the Words of Mathematics" (http://jeff560.tripod.com/mathword.html).

External links

- "Quadratic deviation" (https://www.encyclopediaofmath.org/index.php?title=Quadratic_deviation), *Encyclopedia of Mathematics*, EMS Press, 2001 [1994]
- A simple way to understand Standard Deviation (https://standard-deviation.appspot.com/)
- Standard Deviation – an explanation without maths (http://www.techbookreport.com/tutorials/stddev-30-secs.html)
- The concept of Standard Deviation is shown in this 8-foot-tall (2.4 m) Probability Machine (named Sir Francis) comparing stock market returns to the randomness of the beans dropping through the quincunx pattern. (https://www.youtube.com/watch?v=AUSKTk9ENzg) on YouTube from Index Funds Advisors IFA.com (http://www.ifa.com)

-