[Template:About](/wiki/Template:About" \o "Template:About) In [probability theory](/wiki/Probability_theory) and [statistics](/wiki/Statistics), **variance** is the [expectation](/wiki/Expected_value) of the squared deviation of a [random variable](/wiki/Random_variable) from its [mean](/wiki/Mean), and it informally measures how far a set of (random) numbers are spread out from their mean. The variance has a central role in statistics. It is used in [descriptive statistics](/wiki/Descriptive_statistics), [statistical inference](/wiki/Statistical_inference), [hypothesis testing](/wiki/Hypothesis_testing), [goodness of fit](/wiki/Goodness_of_fit), [Monte Carlo sampling](/wiki/Monte_Carlo_method), amongst many others. This makes it a central quantity in numerous fields such as [physics](/wiki/Physics), [biology](/wiki/Biology), [chemistry](/wiki/Chemistry), [economics](/wiki/Economics), and [finance](/wiki/Finance). The variance is the square of the [standard deviation](/wiki/Standard_deviation), the second [central moment](/wiki/Central_moment) of a distribution, and the [covariance](/wiki/Covariance) of the random variable with itself, and it is often represented by <math>\sigma^2</math> or <math>\operatorname{Var}(X)</math>.

## Contents

* 1 Definition[[edit](/index.php?title=(none)&action=edit&section=1)]
  + 1.1 Continuous random variable[[edit](/index.php?title=(none)&action=edit&section=2)]
  + 1.2 Discrete random variable[[edit](/index.php?title=(none)&action=edit&section=3)]
* 2 Examples[[edit](/index.php?title=(none)&action=edit&section=4)]
  + 2.1 Normal distribution[[edit](/index.php?title=(none)&action=edit&section=5)]
  + 2.2 Exponential distribution[[edit](/index.php?title=(none)&action=edit&section=6)]
  + 2.3 Poisson distribution[[edit](/index.php?title=(none)&action=edit&section=7)]
  + 2.4 Binomial distribution[[edit](/index.php?title=(none)&action=edit&section=8)]
    - 2.4.1 Coin toss[[edit](/index.php?title=(none)&action=edit&section=9)]
  + 2.5 Fair die<!--Singular: die; plural: dice. Don't change-->[[edit](/index.php?title=(none)&action=edit&section=10)]
* 3 Properties[[edit](/index.php?title=(none)&action=edit&section=11)]
  + 3.1 Basic properties[[edit](/index.php?title=(none)&action=edit&section=12)]
  + 3.2 Sum of uncorrelated variables (Bienaymé formula)[[edit](/index.php?title=(none)&action=edit&section=13)]
  + 3.3 Sum of correlated variables[[edit](/index.php?title=(none)&action=edit&section=14)]
  + 3.4 Matrix notation for the variance of a linear combination[[edit](/index.php?title=(none)&action=edit&section=15)]
  + 3.5 Weighted sum of variables[[edit](/index.php?title=(none)&action=edit&section=16)]
  + 3.6 Product of independent variables[[edit](/index.php?title=(none)&action=edit&section=17)]
  + 3.7 Product of correlated variables[[edit](/index.php?title=(none)&action=edit&section=18)]
  + 3.8 Decomposition[[edit](/index.php?title=(none)&action=edit&section=19)]
  + 3.9 Formulae for the variance[[edit](/index.php?title=(none)&action=edit&section=20)]
  + 3.10 Calculation from the CDF[[edit](/index.php?title=(none)&action=edit&section=21)]
  + 3.11 Characteristic property[[edit](/index.php?title=(none)&action=edit&section=22)]
  + 3.12 Units of measurement[[edit](/index.php?title=(none)&action=edit&section=23)]
* 4 Approximating the variance of a function[[edit](/index.php?title=(none)&action=edit&section=24)]
* 5 Population variance and sample variance[[edit](/index.php?title=(none)&action=edit&section=25)]
  + 5.1 Population variance[[edit](/index.php?title=(none)&action=edit&section=26)]
  + 5.2 Sample variance[[edit](/index.php?title=(none)&action=edit&section=27)]
  + 5.3 Distribution of the sample variance[[edit](/index.php?title=(none)&action=edit&section=28)]
  + 5.4 Samuelson's inequality[[edit](/index.php?title=(none)&action=edit&section=29)]
  + 5.5 Relations with the harmonic and arithmetic means[[edit](/index.php?title=(none)&action=edit&section=30)]
* 6 Tests of equality of variances[[edit](/index.php?title=(none)&action=edit&section=31)]
* 7 History[[edit](/index.php?title=(none)&action=edit&section=32)]
* 8 Moment of inertia[[edit](/index.php?title=(none)&action=edit&section=33)]
* 9 Semivariance[[edit](/index.php?title=(none)&action=edit&section=34)]
* 10 Generalizations[[edit](/index.php?title=(none)&action=edit&section=35)]
* 11 See also[[edit](/index.php?title=(none)&action=edit&section=36)]
* 12 Notes[[edit](/index.php?title=(none)&action=edit&section=37)]

## Definition[[edit](/index.php?title=(none)&action=edit&section=1)]

The variance of a random variable <math>X</math> is the [expected value](/wiki/Expected_value) of the squared deviation from the [mean](/wiki/Mean) of <math>X</math>, <math>\mu = \operatorname{E}[X]</math>:

<math> \operatorname{Var}(X) = \operatorname{E}\left[(X - \mu)^2 \right]. </math>

This definition encompasses random variables that are generated by processes that are [discrete](/wiki/Discrete_random_variable), [continuous](/wiki/Continuous_random_variable), [neither](/wiki/Cantor_distribution), or mixed. The variance can also be thought of as the [covariance](/wiki/Covariance) of a random variable with itself:

<math>\operatorname{Var}(X) = \operatorname{Cov}(X, X).</math>

The variance is also equivalent to the second [cumulant](/wiki/Cumulant) of a probability distribution that generates <math>X</math>. The variance is typically designated as <math>\operatorname{Var}(X)</math>, <math>\sigma^2\_X</math>, or simply <math>\sigma^2</math> (pronounced "[sigma](/wiki/Sigma) squared"). The expression for the variance can be expanded:

<math>\begin{align}

\operatorname{Var}(X) &= \operatorname{E}\left[(X - \operatorname{E}[X])^2\right] \\ &= \operatorname{E}\left[X^2 - 2X\operatorname{E}[X] + (\operatorname{E}[X])^2\right] \\ &= \operatorname{E}\left[X^2\right] - 2\operatorname{E}[X]\operatorname{E}[X] + (\operatorname{E}[X])^2 \\ &= \operatorname{E}\left[X^2 \right] - (\operatorname{E}[X])^2 \end{align}</math>

A mnemonic for the above expression is "mean of square minus square of mean". On computational floating point arithmetic, this equation should not be used, because it suffers from [catastrophic cancellation](/wiki/Catastrophic_cancellation) if the two components of the equation are similar in magnitude. There exist [numerically stable alternatives](/wiki/Algorithms_for_calculating_variance).

### Continuous random variable[[edit](/index.php?title=(none)&action=edit&section=2)]

If the random variable <math>X</math> represents samples generated by a [continuous distribution](/wiki/Continuous_distribution) with [probability density function](/wiki/Probability_density_function) <math>f(x)</math>, then the population variance is given by

<math>\operatorname{Var}(X) =\sigma^2 =\int (x-\mu)^2 \, f(x) \, dx\, =\int x^2 \, f(x) \, dx\, - \mu^2</math>

where <math>\mu</math> is the expected value,

<math>\mu = \int x \, f(x) \, dx\, </math>

and where the integrals are [definite integrals](/wiki/Definite_integral) taken for <math>x</math> ranging over the range of <math>X</math>.

If a continuous distribution does not have an expected value, as is the case for the [Cauchy distribution](/wiki/Cauchy_distribution), it does not have a variance either. Many other distributions for which the expected value does exist also do not have a finite variance because the integral in the variance definition diverges. An example is a [Pareto distribution](/wiki/Pareto_distribution) whose [index](/wiki/Pareto_index) <math>k</math> satisfies <math>1 < k \leq 2</math>.

### Discrete random variable[[edit](/index.php?title=(none)&action=edit&section=3)]

If the generator of random variable <math>X</math> is [discrete](/wiki/Discrete_probability_distribution) with [probability mass function](/wiki/Probability_mass_function) <math>x\_1 \mapsto p\_1, x\_2 \mapsto p\_2, \ldots, x\_n \mapsto p\_n</math> then

<math>\operatorname{Var}(X) = \sum\_{i=1}^n p\_i\cdot(x\_i - \mu)^2,</math>

or equivalently

<math>\operatorname{Var}(X) = \sum\_{i=1}^n p\_i x\_i ^2- \mu^2,</math>

where <math>\mu</math> is the expected value, i.e.

<math>\mu = \sum\_{i=1}^n p\_i\cdot x\_i. </math>

(When such a discrete [weighted variance](/wiki/Weighted_variance) is specified by weights whose sum is not 1, then one divides by the sum of the weights.)

The variance of a set of <math>n</math> equally likely values can be written as

<math> \operatorname{Var}(X) = \frac{1}{n} \sum\_{i=1}^n (x\_i - \mu)^2. </math>

where <math>\mu</math> is the expected value, i.e.,

<math>\mu = \frac{1}{n}\sum\_{i=1}^n x\_i </math>

The variance of a set of <math>n</math> equally likely values can be equivalently expressed, without directly referring to the mean, in terms of squared deviations of all points from each other:[[1]](#cite_note-1):<math> \operatorname{Var}(X) = \frac{1}{n^2} \sum\_{i=1}^n \sum\_{j=1}^n \frac{1}{2}(x\_i - x\_j)^2 = \frac{1}{n^2}\sum\_i \sum\_{j>i} (x\_i-x\_j)^2. </math>

## Examples[[edit](/index.php?title=(none)&action=edit&section=4)]

### Normal distribution[[edit](/index.php?title=(none)&action=edit&section=5)]

The [normal distribution](/wiki/Normal_distribution) with parameters μ and σ is a continuous distribution whose [probability density function](/wiki/Probability_density_function) is given by

<math>

f(x) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{ -\frac{(x-\mu)^2}{2\sigma^2} }. </math> In this distribution, E(X) = μ and the variance Var(X) is related with σ via

<math>

\operatorname{Var}(X) = \int\_{-\infty}^\infty \frac{(x - \mu)^2}{\sqrt{2\pi \sigma^2}} e^{ -\frac{(x-\mu)^2}{2\sigma^2} } \, dx = \sigma^2. </math> The role of the normal distribution in the [central limit theorem](/wiki/Central_limit_theorem) is in part responsible for the prevalence of the variance in probability and statistics.

### Exponential distribution[[edit](/index.php?title=(none)&action=edit&section=6)]

The [exponential distribution](/wiki/Exponential_distribution) with parameter <math>\lambda</math> is a continuous distribution whose support is the semi-infinite interval <math>\left[0, \infty\right[</math>. Its [probability density function](/wiki/Probability_density_function) is given by

<math>f(x) = \lambda e^{-\lambda x},\,</math>

and it has expected value <math>\mu = \lambda^{-1}</math>. The variance is equal to

<math> \operatorname{Var}(X) = \int\_0^\infty (x - \lambda^{-1})^2 \, \lambda e^{-\lambda x} dx = \lambda^{-2}.\,</math>

So for an exponentially distributed random variable, <math>\sigma^2 = \mu^2</math>.

### Poisson distribution[[edit](/index.php?title=(none)&action=edit&section=7)]

The [Poisson distribution](/wiki/Poisson_distribution) with parameter <math>\lambda</math> is a discrete distribution for <math>k = 0, 1, 2, \ldots</math>. Its [probability mass function](/wiki/Probability_mass_function) is given by

<math>p(k) = \frac{\lambda^k}{k!} e^{-\lambda},</math>

and it has expected value <math>\mu = \lambda</math>. The variance is equal to

<math> \operatorname{Var}(X) = \sum\_{k=0}^\infty \frac{\lambda^k}{k!} e^{-\lambda} (k-\lambda)^2 = \lambda,</math>

So for a Poisson-distributed random variable, <math>\sigma^2 = \mu</math>.

### Binomial distribution[[edit](/index.php?title=(none)&action=edit&section=8)]

The [binomial distribution](/wiki/Binomial_distribution) with parameters <math>n</math> and <math>p</math> is a discrete distribution for <math>k = 0, 1, 2, \ldots, n</math>. Its [probability mass function](/wiki/Probability_mass_function) is given by

<math>p(k) = {n\choose k}p^k(1-p)^{n-k},</math>

and it has expected value <math>\mu = np</math>. The variance is equal to

<math> \operatorname{Var}(X) = \sum\_{k=0}^{n} {n\choose k}p^k(1-p)^{n-k} (k-np)^2 = np(1-p).</math>

#### Coin toss[[edit](/index.php?title=(none)&action=edit&section=9)]

The binomial distribution with <math>p=0.5</math> describes the probability of getting <math>k</math> heads in <math>n</math> tosses. Thus the expected value of the number of heads is <math alt="n/2">\frac{n}{2}</math>, and the variance is <math alt="n/4">\frac{n}{4}</math>.

### Fair die<!--Singular: die; plural: dice. Don't change-->[[edit](/index.php?title=(none)&action=edit&section=10)]

A six-sided [fair die](/wiki/Dice) can be modelled with a discrete random variable with outcomes 1 through 6, each with equal probability <math display="inline">\frac{1}{6}</math>. The expected value is <math display="inline">\frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3.5</math>. Therefore, the variance can be computed to be

<math>

\begin{align} \sum\_{i=1}^6 \tfrac{1}{6}(i - 3.5)^2 = \tfrac{1}{6}\sum\_{i=1}^6 (i - 3.5)^2 & = \tfrac{1}{6}\left((-2.5)^2{+}(-1.5)^2{+}(-0.5)^2{+}0.5^2{+}1.5^2{+}2.5^2\right) \\ & = \tfrac{1}{6} \cdot 17.50 = \tfrac{35}{12} \approx 2.92. \end{align} </math>

The general formula for the variance of the outcome *X* of a die of n sides is

<math>

\begin{align} \sigma^2=E(X^2)-(E(X))^2 &=\frac{1}{n}\sum\_{i=1}^n i^2-\left(\frac{1}{n}\sum\_{i=1}^n i\right)^2 \\ &=\tfrac 16 (n+1)(2n+1) - \tfrac 14 (n+1)^2\\ &=\frac{ n^2-1 }{12}. \end{align} </math>

## Properties[[edit](/index.php?title=(none)&action=edit&section=11)]

### Basic properties[[edit](/index.php?title=(none)&action=edit&section=12)]

Variance is non-negative because the squares are positive or zero.

<math>\operatorname{Var}(X)\ge 0.</math>

The variance of a constant random variable is zero, and if the variance of a variable in a [data set](/wiki/Data_set) is 0, then all the entries have the same value.

<math>P(X=a) = 1\Leftrightarrow \operatorname{Var}(X)= 0.</math>

Variance is [invariant](/wiki/Invariant_(mathematics)) with respect to changes in a [location parameter](/wiki/Location_parameter). That is, if a constant is added to all values of the variable, the variance is unchanged.

<math>\operatorname{Var}(X+a)=\operatorname{Var}(X).</math>

If all values are scaled by a constant, the variance is scaled by the square of that constant.

<math>\operatorname{Var}(aX)=a^2\operatorname{Var}(X).</math>

The variance of a sum of two random variables is given by:

<math>\operatorname{Var}(aX+bY)=a^2\operatorname{Var}(X)+b^2\operatorname{Var}(Y)+2ab\, \operatorname{Cov}(X,Y),</math>

<math>\operatorname{Var}(aX-bY)=a^2\operatorname{Var}(X)+b^2\operatorname{Var}(Y)-2ab\, \operatorname{Cov}(X,Y),</math>

where [Template:Math](/wiki/Template:Math) is the [covariance](/wiki/Covariance). In general we have for the sum of <math>N</math> random variables <math>\{X\_1,\dots,X\_N\}</math>:

<math>\operatorname{Var}\left(\sum\_{i=1}^N X\_i\right)=\sum\_{i,j=1}^N\operatorname{Cov}(X\_i,X\_j)=\sum\_{i=1}^N\operatorname{Var}(X\_i)+\sum\_{i\ne j}\operatorname{Cov}(X\_i,X\_j).</math>

These results lead to the variance of a [linear combination](/wiki/Linear_combination) as:

<math>

\begin{align} \operatorname{Var}\left( \sum\_{i=1}^N a\_iX\_i\right) &=\sum\_{i,j=1}^{N} a\_ia\_j\operatorname{Cov}(X\_i,X\_j) \\ &=\sum\_{i=1}^N a\_i^2\operatorname{Var}(X\_i)+\sum\_{i\not=j}a\_ia\_j\operatorname{Cov}(X\_i,X\_j)\\ & =\sum\_{i=1}^N a\_i^2\operatorname{Var}(X\_i)+2\sum\_{1\le i<j\le N}a\_ia\_j\operatorname{Cov}(X\_i,X\_j). \end{align} </math>

If the random variables <math>X\_1,\dots,X\_N</math> are such that

<math>\operatorname{Cov}(X\_i,X\_j)=0\ ,\ \forall\ (i\ne j) ,</math>

they are said to be [uncorrelated](/wiki/Covariance#Definition). It follows immediately from the expression given earlier that if the random variables <math>X\_1,\dots,X\_N</math> are uncorrelated, then the variance of their sum is equal to the sum of their variances, or, expressed symbolically:

<math>\operatorname{Var}\left(\sum\_{i=1}^N X\_i\right)=\sum\_{i=1}^N\operatorname{Var}(X\_i).</math>

Since [independent random variables are always uncorrelated](/wiki/Covariance#Uncorrelatedness_and_independence), the equation above holds in particular when the random variables <math>X\_1,\dots,X\_n</math> are independent. Thus independence is sufficient but not necessary for the variance of the sum to equal the sum of the variances.

### Sum of uncorrelated variables (Bienaymé formula)[[edit](/index.php?title=(none)&action=edit&section=13)]

[Template:See also](/wiki/Template:See_also) One reason for the use of the variance in preference to other measures of dispersion is that the variance of the sum (or the difference) of [uncorrelated](/wiki/Uncorrelated) random variables is the sum of their variances:

<math>\operatorname{Var}\left(\sum\_{i=1}^n X\_i\right) = \sum\_{i=1}^n \operatorname{Var}(X\_i).</math>

This statement is called the [Bienaymé](/wiki/Irénée-Jules_Bienaymé) formula[[2]](#cite_note-2) and was discovered in 1853.[Template:Citation needed](/wiki/Template:Citation_needed) It is often made with the stronger condition that the variables are [independent](/wiki/Statistical_independence), but being uncorrelated suffices. So if all the variables have the same variance σ2, then, since division by *n* is a linear transformation, this formula immediately implies that the variance of their mean is

<math>\operatorname{Var}\left(\overline{X}\right) = \operatorname{Var}\left(\frac {1} {n}\sum\_{i=1}^n X\_i\right) = \frac {1} {n^2}\sum\_{i=1}^n \operatorname{Var}\left(X\_i\right) = \frac {\sigma^2} {n}.</math>

That is, the variance of the mean decreases when *n* increases. This formula for the variance of the mean is used in the definition of the [standard error](/wiki/Standard_error_(statistics)) of the sample mean, which is used in the [central limit theorem](/wiki/Central_limit_theorem).

### Sum of correlated variables[[edit](/index.php?title=(none)&action=edit&section=14)]

In general, if the variables are [correlated](/wiki/Correlated), then the variance of their sum is the sum of their [covariances](/wiki/Covariance):

<math>\operatorname{Var}\left(\sum\_{i=1}^n X\_i\right) = \sum\_{i=1}^n \sum\_{j=1}^n \operatorname{Cov}(X\_i, X\_j) = \sum\_{i=1}^n \operatorname{Var}(X\_i) + 2\sum\_{1\le i<j\le n}\operatorname{Cov}(X\_i,X\_j).</math>

(Note: The second equality comes from the fact that [Template:Math](/wiki/Template:Math).)

Here [Template:Math](/wiki/Template:Math) is the [covariance](/wiki/Covariance), which is zero for independent random variables (if it exists). The formula states that the variance of a sum is equal to the sum of all elements in the covariance matrix of the components. This formula is used in the theory of [Cronbach's alpha](/wiki/Cronbach's_alpha) in [classical test theory](/wiki/Classical_test_theory).

So if the variables have equal variance *σ*2 and the average correlation of distinct variables is *ρ*, then the variance of their mean is

<math>\operatorname{Var}(\overline{X}) = \frac {\sigma^2} {n} + \frac {n-1} {n} \rho \sigma^2.</math>

This implies that the variance of the mean increases with the average of the correlations. In other words, additional correlated observations are not as effective as additional independent observations at reducing the [uncertainty of the mean](/wiki/Standard_error). Moreover, if the variables have unit variance, for example if they are standardized, then this simplifies to

<math>\operatorname{Var}(\overline{X}) = \frac {1} {n} + \frac {n-1} {n} \rho.</math>

This formula is used in the [Spearman–Brown prediction formula](/wiki/Spearman–Brown_prediction_formula) of classical test theory. This converges to *ρ* if *n* goes to infinity, provided that the average correlation remains constant or converges too. So for the variance of the mean of standardized variables with equal correlations or converging average correlation we have

<math> \lim\_{n \to \infty} \operatorname{Var}(\overline{X}) = \rho.</math>

Therefore, the variance of the mean of a large number of standardized variables is approximately equal to their average correlation. This makes clear that the sample mean of correlated variables does not generally converge to the population mean, even though the [Law of large numbers](/wiki/Law_of_large_numbers) states that the sample mean will converge for independent variables.

### Matrix notation for the variance of a linear combination[[edit](/index.php?title=(none)&action=edit&section=15)]

Define <math>X</math> as a column vector of <math>n</math> random variables <math>X\_1, \ldots,X\_n</math>, and <math>c</math> as a column vector of <math>n</math> scalars <math>c\_1, \ldots,c\_n</math>. Therefore, <math>c^T X</math> is a [linear combination](/wiki/Linear_combination) of these random variables, where <math>c^T</math> denotes the [transpose](/wiki/Transpose) of <math>c</math>. Also let <math>\Sigma</math> be the [covariance matrix](/wiki/Covariance_matrix) of <math>X</math>. The variance of <math>c^TX</math> is then given by:[[3]](#cite_note-3)

<math>\operatorname{Var}(c^T X) = c^T \Sigma c .</math>

### Weighted sum of variables[[edit](/index.php?title=(none)&action=edit&section=16)]

The scaling property and the Bienaymé formula, along with the property of the [covariance](/wiki/Covariance) [Template:Math](/wiki/Template:Math) jointly imply that

<math>\operatorname{Var}(aX \pm bY) =a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y) \pm 2ab\, \operatorname{Cov}(X, Y).</math>

This implies that in a weighted sum of variables, the variable with the largest weight will have a disproportionally large weight in the variance of the total. For example, if *X* and *Y* are uncorrelated and the weight of *X* is two times the weight of *Y*, then the weight of the variance of *X* will be four times the weight of the variance of *Y*.

The expression above can be extended to a weighted sum of multiple variables:

<math>\operatorname{Var}\left(\sum\_{i}^n a\_iX\_i\right) = \sum\_{i=1}^na\_i^2 \operatorname{Var}(X\_i) + 2\sum\_{1\le i}\sum\_{<j\le n}a\_ia\_j\operatorname{Cov}(X\_i,X\_j)</math>

### Product of independent variables[[edit](/index.php?title=(none)&action=edit&section=17)]

If two variables X and Y are [independent](/wiki/Independence_(probability_theory)), the variance of their product is given by[[4]](#cite_note-4):<math> \begin{align} \operatorname{Var}(XY) &= [E(X)]^2 \operatorname{Var}(Y) + [E(Y)]^2 \operatorname{Var}(X) + \operatorname{Var}(X)\operatorname{Var}(Y). \end{align} </math>

Equivalently, using the basic properties of expectation, it is given by

<math>

\operatorname{Var}(XY) = E(X^2) E(Y^2) - [E(X)]^2 [E(Y)]^2. </math>

### Product of correlated variables[[edit](/index.php?title=(none)&action=edit&section=18)]

In general, if two variables are [correlated](/wiki/Correlated), the variance of their product is given by:

<math>

\begin{align} \operatorname{Var}(XY) &= E[X^2 Y^2 ]-[E(XY)]^2 \\ &= \operatorname{Cov}(X^2,Y^2 )+E(X^2)E(Y^2) - [E(XY)]^2 \\ &= \operatorname{Cov}(X^2, Y^2) +(\operatorname{Var}(X)+[E(X)]^2 )(\operatorname{Var}(Y)+[E(Y)]^2 )-[\operatorname{Cov}(X,Y)+E(X)E(Y)]^2 \end{align}</math>

### Decomposition[[edit](/index.php?title=(none)&action=edit&section=19)]

The general formula for variance decomposition or the [law of total variance](/wiki/Law_of_total_variance) is: If <math>X</math> and <math>Y</math> are two random variables, and the variance of <math>X</math> exists, then

<math>\operatorname{Var}[X]=\operatorname{E}\_Y(\operatorname{Var}[X\mid Y])+\operatorname{Var}\_Y(\operatorname{E}[X\mid Y]).\,</math>

where <math>\operatorname E(X|Y)</math> is the [conditional expectation](/wiki/Conditional_expectation) of <math>X</math> given <math>Y</math>, and <math>\operatorname{Var}(X|Y)</math> is the [conditional variance](/wiki/Conditional_variance) of <math>X</math> given <math>Y</math>. (A more intuitive explanation is that given a particular value of <math>Y</math>, then <math>X</math> follows a distribution with mean <math>\operatorname E(X|Y)</math> and variance <math>\operatorname{Var}(X|Y)</math>. As <math>\operatorname E(X|Y)</math> is a function of the variable <math>Y</math>, the outer expectation or variance is taken with respect to Y. The above formula tells how to find <math>\operatorname{Var}(X)</math> based on the distributions of these two quantities when <math>Y</math> is allowed to vary. A similar formula is applied in [analysis of variance](/wiki/Analysis_of_variance), where the corresponding formula is

<math>\mathit{MS}\_\text{total} = \mathit{MS}\_\text{between} + \mathit{MS}\_\text{within};</math>

here <math>\mathit{MS}</math> refers to the Mean of the Squares. In [linear regression](/wiki/Linear_regression) analysis the corresponding formula is

<math>\mathit{MS}\_\text{total} = \mathit{MS}\_\text{regression} + \mathit{MS}\_\text{residual}.</math>

This can also be derived from the additivity of variances, since the total (observed) score is the sum of the predicted score and the error score, where the latter two are uncorrelated.

Similar decompositions are possible for the sum of squared deviations (sum of squares, <math>\mathit{SS}</math>):

<math>\mathit{SS}\_\text{total} = \mathit{SS}\_\text{between} + \mathit{SS}\_\text{within},</math>

<math>\mathit{SS}\_\text{total} = \mathit{SS}\_\text{regression} + \mathit{SS}\_\text{residual}.</math>

### Formulae for the variance[[edit](/index.php?title=(none)&action=edit&section=20)]

[Template:Main](/wiki/Template:Main) A formula often used for deriving the variance of a theoretical distribution is as follows:

<math> \operatorname{Var}(X) =\operatorname{E}(X^2) - (\operatorname{E}(X))^2. </math>

This will be useful when it is possible to derive formulae for the expected value and for the expected value of the square.

This formula is also sometimes used in connection with the sample variance. While useful for hand calculations, it is not advised for computer calculations as it suffers from [catastrophic cancellation](/wiki/Catastrophic_cancellation) if the two components of the equation are similar in magnitude and floating point arithmetic is used. This is discussed in the article [Algorithms for calculating variance](/wiki/Algorithms_for_calculating_variance).

### Calculation from the CDF[[edit](/index.php?title=(none)&action=edit&section=21)]

The population variance for a non-negative random variable can be expressed in terms of the [cumulative distribution function](/wiki/Cumulative_distribution_function) *F* using

<math>

2\int\_0^\infty u( 1-F(u))\,du - \Big(\int\_0^\infty 1-F(u)\,du\Big)^2. </math>

This expression can be used to calculate the variance in situations where the CDF, but not the [density](/wiki/Probability_density_function), can be conveniently expressed.

### Characteristic property[[edit](/index.php?title=(none)&action=edit&section=22)]

The second [moment](/wiki/Moment_(mathematics)) of a random variable attains the minimum value when taken around the first moment (i.e., mean) of the random variable, i.e. <math>\mathrm{argmin}\_m\,\mathrm{E}((X - m)^2) = \mathrm{E}(X)\,</math>. Conversely, if a continuous function <math>\varphi</math> satisfies <math>\mathrm{argmin}\_m\,\mathrm{E}(\varphi(X - m)) = \mathrm{E}(X)\,</math> for all random variables *X*, then it is necessarily of the form <math>\varphi(x) = a x^2 + b</math>, where [Template:Nowrap](/wiki/Template:Nowrap). This also holds in the multidimensional case.[[5]](#cite_note-5)

### Units of measurement[[edit](/index.php?title=(none)&action=edit&section=23)]

Unlike expected absolute deviation, the variance of a variable has units that are the square of the units of the variable itself. For example, a variable measured in meters will have a variance measured in meters squared. For this reason, describing data sets via their [standard deviation](/wiki/Standard_deviation) or [root mean square deviation](/wiki/Root_mean_square_deviation) is often preferred over using the variance. In the dice example the standard deviation is √2.9 ≈ 1.7, slightly larger than the expected absolute deviation of 1.5.

The standard deviation and the expected absolute deviation can both be used as an indicator of the "spread" of a distribution. The standard deviation is more amenable to algebraic manipulation than the expected absolute deviation, and, together with variance and its generalization [covariance](/wiki/Covariance), is used frequently in theoretical statistics; however the expected absolute deviation tends to be more [robust](/wiki/Robust_statistics) as it is less sensitive to [outliers](/wiki/Outlier) arising from [measurement anomalies](/wiki/Measurement_error) or an unduly [heavy-tailed distribution](/wiki/Heavy-tailed_distribution).

## Approximating the variance of a function[[edit](/index.php?title=(none)&action=edit&section=24)]

The [delta method](/wiki/Delta_method) uses second-order [Taylor expansions](/wiki/Taylor_expansion) to approximate the variance of a function of one or more random variables: see [Taylor expansions for the moments of functions of random variables](/wiki/Taylor_expansions_for_the_moments_of_functions_of_random_variables). For example, the approximate variance of a function of one variable is given by

<math>\operatorname{Var}\left[f(X)\right]\approx \left(f'(\operatorname{E}\left[X\right])\right)^2\operatorname{Var}\left[X\right]</math>

provided that *f* is twice differentiable and that the mean and variance of *X* are finite.

## Population variance and sample variance[[edit](/index.php?title=(none)&action=edit&section=25)]

[Template:Anchor](/wiki/Template:Anchor) [Template:See also](/wiki/Template:See_also) [Template:Move section portions](/wiki/Template:Move_section_portions)

Real-world observations such as the measurements of yesterday's rain throughout the day typically cannot be complete sets of all possible observations that could be made. As such, the variance calculated from the finite set will in general not match the variance that would have been calculated from the full population of possible observations. This means that one [estimates](/wiki/Estimation_theory) the mean and variance that would have been calculated from an omniscient set of observations by using an [estimator](/wiki/Estimator) equation. The estimator is a function of the [sample](/wiki/Sample_(statistics)) of *n* [observations](/wiki/Observations) drawn without observational bias from the whole [population](/wiki/Population) of potential observations. In this example that sample would be the set of actual measurements of yesterday's rainfall from available rain gauges within the geography of interest.

The simplest estimators for population mean and population variance are simply the mean and variance of the sample, the **sample mean** and **(uncorrected) sample variance** – these are [consistent estimators](/wiki/Consistent_estimator) (they converge to the correct value as the number of samples increases), but can be improved. Estimating the population variance by taking the sample's variance is close to optimal in general, but can be improved in two ways. Most simply, the sample variance is computed as an average of [squared deviations](/wiki/Squared_deviations) about the (sample) mean, by dividing by *n.* However, using values other than *n* improves the estimator in various ways. Four common values for the denominator are *n,* *n* − 1, *n* + 1, and *n* − 1.5: *n* is the simplest (population variance of the sample), *n* − 1 eliminates bias, *n* + 1 minimizes [mean squared error](/wiki/Mean_squared_error) for the normal distribution, and *n* − 1.5 mostly eliminates bias in [unbiased estimation of standard deviation](/wiki/Unbiased_estimation_of_standard_deviation) for the normal distribution.

Firstly, if the omniscient mean is unknown (and is computed as the sample mean), then the sample variance is a [biased estimator](/wiki/Biased_estimator): it underestimates the variance by a factor of (*n* − 1) / *n*; correcting by this factor (dividing by *n* − 1 instead of *n*) is called [Bessel's correction](/wiki/Bessel's_correction). The resulting estimator is unbiased, and is called the **(corrected) sample variance** or **unbiased sample variance**. For example, when *n* = 1 the variance of a single observation about the sample mean (itself) is obviously zero regardless of the population variance. If the mean is determined in some other way than from the same samples used to estimate the variance then this bias does not arise and the variance can safely be estimated as that of the samples about the (independently known) mean.

Secondly, the sample variance does not generally minimize [mean squared error](/wiki/Mean_squared_error) between sample variance and population variance. Correcting for bias often makes this worse: one can always choose a scale factor that performs better than the corrected sample variance, though the optimal scale factor depends on the [excess kurtosis](/wiki/Excess_kurtosis) of the population (see [mean squared error: variance](/wiki/Mean_squared_error#Variance)), and introduces bias. This always consists of scaling down the unbiased estimator (dividing by a number larger than *n* − 1), and is a simple example of a [shrinkage estimator](/wiki/Shrinkage_estimator): one "shrinks" the unbiased estimator towards zero. For the normal distribution, dividing by *n* + 1 (instead of *n* − 1 or *n*) minimizes mean squared error. The resulting estimator is biased, however, and is known as the **biased sample variation**.

### Population variance[[edit](/index.php?title=(none)&action=edit&section=26)]

In general, the ***population variance*** of a *finite* [population](/wiki/Statistical_population) of size *N* with values *xi* is given by

<math> \sigma^2 = \frac 1N \sum\_{i=1}^N \left(x\_i - \mu \right)^2 = \frac 1N \sum\_{i=1}^N \left(x\_i^2 - 2\mu x\_i + \mu^2 \right) = \left(\frac 1N \sum\_{i=1}^N x\_i^2\right) - 2\mu \left(\frac 1N \sum\_{i=1}^N x\_i\right) + \mu^2 = \left(\frac 1N \sum\_{i=1}^N x\_i^2\right) - \mu^2</math>

where the population mean is

<math> \mu = \frac 1N \sum\_{i=1}^N x\_i </math>.

Also,

<math> \sigma^2 = \frac {1} {N^2}\sum\_{i<j}\left( x\_i-x\_j \right)^2 = \frac{1}{2N^2} \sum\_{i, j=1}^N\left( x\_i-x\_j \right)^2</math>

because

<math> \frac{1}{2N^2} \sum\_{i, j=1}^N\left( x\_i-x\_j \right)^2 = \frac{1}{2N^2} \sum\_{i, j=1}^N\left( x\_i^2 - 2x\_ix\_j + x\_j^2 \right) = \frac{1}{2N} \sum\_{j=1}^N\left(\frac 1N \sum\_{i=1}^N x\_i^2\right) - \left(\frac 1N \sum\_{i=1}^N x\_i^2\right)\left(\frac 1N \sum\_{j=1}^N x\_j^2\right) + \frac{1}{2N} \sum\_{i=1}^N\left(\frac 1N \sum\_{j=1}^N x\_j^2\right) = \frac{1}{2} \left( \sigma^2 + \mu^2 \right) - \mu^2 + \frac{1}{2} \left( \sigma^2 + \mu^2 \right)</math>

The population variance therefore matches the variance of the generating probability distribution. In this sense, the concept of population can be extended to continuous random variables with infinite populations.

### Sample variance[[edit](/index.php?title=(none)&action=edit&section=27)]

In many practical situations, the true variance of a population is not known *a priori* and must be computed somehow. When dealing with extremely large populations, it is not possible to count every object in the population, so the computation must be performed on a [sample](/wiki/Sample_(statistics)) of the population.[[6]](#cite_note-6) Sample variance can also be applied to the estimation of the variance of a continuous distribution from a sample of that distribution.

We take a [sample with replacement](/wiki/Statistical_sample) of *n* values *y*1, ..., *yn* from the population, where *n* < *N*, and estimate the variance on the basis of this sample.[[7]](#cite_note-7) Directly taking the variance of the sample data gives the average of the [squared deviations](/wiki/Squared_deviations):

<math>\sigma\_y^2 = \frac 1n \sum\_{i=1}^n \left(y\_i - \overline{y} \right)^2 =\left(\frac 1n \sum\_{i=1}^n y\_i^2\right) - \overline{y}^2 = \frac {1} {n^2}\sum\_{i<j}\left( y\_i-y\_j \right)^2.</math>

Here, <math>\overline{y}</math> denotes the [sample mean](/wiki/Sample_mean):

<math>\overline{y}=\frac 1n \sum\_{i=1}^n y\_i .</math>

Since the *yi* are selected randomly, both <math>\scriptstyle\overline{y}</math> and <math>\scriptstyle\sigma\_y^2</math> are random variables. Their expected values can be evaluated by summing over the ensemble of all possible samples {*yi*} from the population. For <math>\scriptstyle\sigma\_y^2</math> this gives:

<math>

\begin{align} E[\sigma\_y^2] & = E\left[ \frac 1n \sum\_{i=1}^n \left(y\_i - \frac 1n \sum\_{j=1}^n y\_j \right)^2 \right] \\ & = \frac 1n \sum\_{i=1}^n E\left[ y\_i^2 - \frac 2n y\_i \sum\_{j=1}^n y\_j + \frac{1}{n^2} \sum\_{j=1}^n y\_j \sum\_{k=1}^n y\_k \right] \\ & = \frac 1n \sum\_{i=1}^n \left[ \frac{n-2}{n} E[y\_i^2] - \frac 2n \sum\_{j \neq i} E[y\_i y\_j] + \frac{1}{n^2} \sum\_{j=1}^n \sum\_{k \neq j}^n E[y\_j y\_k] +\frac{1}{n^2} \sum\_{j=1}^n E[y\_j^2] \right] \\ & = \frac 1n \sum\_{i=1}^n \left[ \frac{n-2}{n} (\sigma^2+\mu^2) - \frac 2n (n-1) \mu^2 + \frac{1}{n^2} n (n-1) \mu^2 + \frac 1n (\sigma^2+\mu^2) \right] \\ & = \frac{n-1}{n} \sigma^2. \end{align} </math>

Hence <math>\scriptstyle\sigma\_y^2</math> gives an estimate of the population variance that is biased by a factor of <math>\frac{n-1}{n}</math>. For this reason, <math>\scriptstyle\sigma\_y^2</math> is referred to as the *biased sample variance*. Correcting for this bias yields the *unbiased sample variance*:

<math>s^2 = \frac{n}{n-1} \sigma\_y^2 = \frac{n}{n-1} \left( \frac{1}{n} \sum\_{i=1}^n \left(y\_i - \overline{y} \right)^2 \right) = \frac{1}{n-1} \sum\_{i=1}^n \left(y\_i - \overline{y} \right)^2 </math>

Either estimator may be simply referred to as the *sample variance* when the version can be determined by context. The same proof is also applicable for samples taken from a continuous probability distribution.

The use of the term *n* − 1 is called [Bessel's correction](/wiki/Bessel's_correction), and it is also used in [sample covariance](/wiki/Sample_covariance) and the [sample standard deviation](/wiki/Sample_standard_deviation) (the square root of variance). The square root is a [concave function](/wiki/Concave_function) and thus introduces negative bias (by [Jensen's inequality](/wiki/Jensen's_inequality)), which depends on the distribution, and thus the corrected sample standard deviation (using Bessel's correction) is biased. The [unbiased estimation of standard deviation](/wiki/Unbiased_estimation_of_standard_deviation) is a technically involved problem, though for the normal distribution using the term *n* − 1.5 yields an almost unbiased estimator.

The unbiased sample variance is a [U-statistic](/wiki/U-statistic) for the function *ƒ*(*y*1, *y*2) = (*y*1 − *y*2)2/2, meaning that it is obtained by averaging a 2-sample statistic over 2-element subsets of the population.

### Distribution of the sample variance[[edit](/index.php?title=(none)&action=edit&section=28)]

[Template:Multiple image](/wiki/Template:Multiple_image) Being a function of [random variables](/wiki/Random_variable), the sample variance is itself a random variable, and it is natural to study its distribution. In the case that *yi* are independent observations from a [normal distribution](/wiki/Normal_distribution), [Cochran's theorem](/wiki/Cochran's_theorem) shows that *s*2 follows a scaled [chi-squared distribution](/wiki/Chi-squared_distribution):[[8]](#cite_note-8):<math> (n-1)\frac{s^2}{\sigma^2}\sim\chi^2\_{n-1}. </math>

As a direct consequence, it follows that

<math>

\operatorname{E}(s^2)=\operatorname{E}\left(\frac{\sigma^2}{n-1} \chi^2\_{n-1}\right)=\sigma^2 , </math>

and[[9]](#cite_note-9)

<math>

\operatorname{Var}[s^2] =\operatorname{Var}\left(\frac{\sigma^2}{n-1} \chi^2\_{n-1}\right)=\frac{\sigma^4}{(n-1)^2}\operatorname{Var}\left( \chi^2\_{n-1}\right)=\frac{2\sigma^4 }{n-1}.

</math>

If the *yi* are independent and identically distributed, but not necessarily normally distributed, then[[10]](#cite_note-10)

<math>

\operatorname{E}[s^2] = \sigma^2, \quad

\operatorname{Var}[s^2] = \frac{\sigma^4}{n} \left ((\kappa-1)+\frac{2}{n-1} \right) = \frac{1}{n} \left(\mu\_4 - \frac{n-3}{n-1}\sigma^4\right),

</math>

where *κ* is the [kurtosis](/wiki/Kurtosis) of the distribution and *μ4* is the fourth [central moment](/wiki/Central_moment).

If the conditions of the [law of large numbers](/wiki/Law_of_large_numbers) hold for the squared observations, *s*2 is a [consistent estimator](/wiki/Consistent_estimator) of *σ*2. One can see indeed that the variance of the estimator tends asymptotically to zero.

### Samuelson's inequality[[edit](/index.php?title=(none)&action=edit&section=29)]

[Samuelson's inequality](/wiki/Samuelson's_inequality) is a result that states bounds on the values that individual observations in a sample can take, given that the sample mean and (biased) variance have been calculated.[[11]](#cite_note-11) Values must lie within the limits <math>\bar y \pm \sigma\_y (n-1)^{1/2}.</math>

### Relations with the harmonic and arithmetic means[[edit](/index.php?title=(none)&action=edit&section=30)]

It has been shown[[12]](#cite_note-12) that for a sample {*yi*} of real numbers,

<math> \sigma\_y^2 \le 2y\_{\max} (A - H), </math>

where *y*max is the maximum of the sample, *A* is the arithmetic mean, *H* is the [harmonic mean](/wiki/Harmonic_mean) of the sample and <math>\sigma\_y^2</math> is the (biased) variance of the sample.

This bound has been improved, and it is known that variance is bounded by

<math> \sigma\_y^2 \le \frac{y\_{\max} (A - H)(y\_\max - A)}{y\_\max - H}, </math>

<math> \sigma\_y^2 \ge \frac{y\_{\min} (A - H)(A - y\_\min)}{H - y\_\min}, </math>

where *y*min is the minimum of the sample.<ref name=Sharma2008>[Template:Cite journal](/wiki/Template:Cite_journal)</ref>

## Tests of equality of variances[[edit](/index.php?title=(none)&action=edit&section=31)]

Testing for the equality of two or more variances is difficult. The [F test](/wiki/F_test) and [chi square tests](/wiki/Chi_square_test) are both adversely affected by non-normality and are not recommended for this purpose.

Several non parametric tests have been proposed: these include the Barton–David–Ansari–Freund–Siegel–Tukey test, the [Capon test](/wiki/Capon_test), [Mood test](/wiki/Mood_test), the [Klotz test](/wiki/Klotz_test) and the [Sukhatme test](/wiki/Sukhatme_test). The Sukhatme test applies to two variances and requires that both [medians](/wiki/Median) be known and equal to zero. The Mood, Klotz, Capon and Barton–David–Ansari–Freund–Siegel–Tukey tests also apply to two variances. They allow the median to be unknown but do require that the two medians are equal.

The [Lehmann test](/wiki/Lehmann_test) is a parametric test of two variances. Of this test there are several variants known. Other tests of the equality of variances include the [Box test](/wiki/Box_test), the [Box–Anderson test](/wiki/Box–Anderson_test) and the [Moses test](/wiki/Moses_test).

Resampling methods, which include the [bootstrap](/wiki/Bootstrapping_(statistics)) and the [jackknife](/wiki/Resampling_(statistics)), may be used to test the equality of variances.

## History[[edit](/index.php?title=(none)&action=edit&section=32)]

The term *variance* was first introduced by [Ronald Fisher](/wiki/Ronald_Fisher) in his 1918 paper [*The Correlation Between Relatives on the Supposition of Mendelian Inheritance*](/wiki/The_Correlation_Between_Relatives_on_the_Supposition_of_Mendelian_Inheritance):[[13]](#cite_note-13)

The great body of available statistics show us that the deviations of a [human measurement](/wiki/Biometry) from its mean follow very closely the [Normal Law of Errors](/wiki/Normal_distribution), and, therefore, that the variability may be uniformly measured by the [standard deviation](/wiki/Standard_deviation) corresponding to the [square root](/wiki/Square_root) of the [mean square error](/wiki/Mean_square_error). When there are two independent causes of variability capable of producing in an otherwise uniform population distributions with standard deviations <math>\sigma\_1</math> and <math>\sigma\_2</math>, it is found that the distribution, when both causes act together, has a standard deviation <math>\sqrt{\sigma\_1^2 + \sigma\_2^2}</math>. It is therefore desirable in analysing the causes of variability to deal with the square of the standard deviation as the measure of variability. We shall term this quantity the Variance...

[thumb|Geometric visualisation of the variance of an arbitrary distribution (2, 4, 4, 4, 5, 5, 7, 9):   
1. A frequency distribution is constructed.   
2. The centroid of the distribution gives its mean.   
3. A square with sides equal to the difference of each value from the mean is formed for each value.   
4. Arranging the squares into a rectangle with one side equal to the number of values, *n*, results in the other side being the distribution's variance, *σ*².](/wiki/File:variance_visualisation.svg)

## Moment of inertia[[edit](/index.php?title=(none)&action=edit&section=33)]

[Template:See also](/wiki/Template:See_also) The variance of a probability distribution is analogous to the [moment of inertia](/wiki/Moment_of_inertia) in [classical mechanics](/wiki/Classical_mechanics) of a corresponding mass distribution along a line, with respect to rotation about its center of mass.[Template:Citation needed](/wiki/Template:Citation_needed) It is because of this analogy that such things as the variance are called [*moments*](/wiki/Moment_(mathematics)) of [probability distributions](/wiki/Probability_distribution).[Template:Citation needed](/wiki/Template:Citation_needed) The covariance matrix is related to the [moment of inertia tensor](/wiki/Moment_of_inertia_tensor) for multivariate distributions. The moment of inertia of a cloud of *n* points with a covariance matrix of <math>\Sigma</math> is given by[Template:Citation needed](/wiki/Template:Citation_needed)

<math>I=n (\mathbf{1}\_{3\times 3} \operatorname{tr}(\Sigma) - \Sigma).</math>

This difference between moment of inertia in physics and in statistics is clear for points that are gathered along a line. Suppose many points are close to the *x* axis and distributed along it. The covariance matrix might look like

<math>\Sigma=\begin{bmatrix}10 & 0 & 0\\0 & 0.1 & 0 \\ 0 & 0 & 0.1\end{bmatrix}.</math>

That is, there is the most variance in the *x* direction. Physicists would consider this to have a low moment *about* the *x* axis so the moment-of-inertia tensor is

<math>I=n\begin{bmatrix}0.2 & 0 & 0\\0 & 10.1 & 0 \\ 0 & 0 & 10.1\end{bmatrix}.</math>

## Semivariance[[edit](/index.php?title=(none)&action=edit&section=34)]

The *semivariance* is calculated in the same manner as the variance but only those observations that fall below the mean are included in the calculation. It is sometimes described as a measure of [downside risk](/wiki/Downside_risk) in an [investments](/wiki/Investment#In_finance) context. For skewed distributions, the semivariance can provide additional information that a variance does not[Template:Citation needed](/wiki/Template:Citation_needed).

## Generalizations[[edit](/index.php?title=(none)&action=edit&section=35)]

[Template:Merge](/wiki/Template:Merge)

If <math>x</math> is a scalar [complex](/wiki/Complex_number)-valued random variable, with values in <math>\mathbb{C}</math>, then its variance is <math>\operatorname{E}((x - \mu)(x - \mu)^\*)</math>, where <math>x^\*</math> is the [complex conjugate](/wiki/Complex_conjugate) of <math>x</math>. This variance is a real scalar.

If <math>X</math> is a [vector](/wiki/Vector_space)-valued random variable, with values in <math>\mathbb{R}^n</math>, and thought of as a column vector, then the natural generalization of variance is <math>\operatorname{E}((X - \mu)(X - \mu)^{\operatorname{T}})</math>, where <math>\mu = \operatorname{E}(X)</math> and <math>X^{\operatorname{T}}</math> is the transpose of <math>X</math>, and so is a row vector. The result is a [positive semi-definite square matrix](/wiki/Positive_definite_matrix), commonly referred to as the [variance-covariance matrix](/wiki/Variance-covariance_matrix).

If <math>X</math> is a vector- and complex-valued random variable, with values in <math>\mathbb{C}^n</math>, then the generalization of its variance is <math>\operatorname{E}((X - \mu)(X - \mu)^\dagger)</math>, where <math>X^\dagger</math> is the [conjugate transpose](/wiki/Conjugate_transpose) of <math>X</math>. This matrix is also positive semi-definite and square.

## See also[[edit](/index.php?title=(none)&action=edit&section=36)]

[Template:Portal](/wiki/Template:Portal) [Template:Wiktionary](/wiki/Template:Wiktionary) [Template:Colbegin](/wiki/Template:Colbegin)

* [Algorithms for calculating variance](/wiki/Algorithms_for_calculating_variance)
* [Average absolute deviation](/wiki/Average_absolute_deviation)
* [Bhatia–Davis inequality](/wiki/Bhatia–Davis_inequality)
* [Common-method variance](/wiki/Common-method_variance)
* [Correlation](/wiki/Correlation)
* [Covariance](/wiki/Covariance)
* [Chebyshev's inequality](/wiki/Chebyshev's_inequality)
* [Distance variance](/wiki/Distance_variance)
* [Estimation of covariance matrices](/wiki/Estimation_of_covariance_matrices)
* [Explained variance](/wiki/Explained_variance)
* [Homoscedasticity](/wiki/Homoscedasticity)
* [Mean absolute error](/wiki/Mean_absolute_error)
* [Mean difference](/wiki/Mean_absolute_difference)
* [Mean preserving spread](/wiki/Mean_preserving_spread)
* [Pooled variance](/wiki/Pooled_variance) (also known as combined, composite, or overall variance)
* [Popoviciu's inequality on variances](/wiki/Popoviciu's_inequality_on_variances)
* [Qualitative variation](/wiki/Qualitative_variation)
* [Sample mean and covariance](/wiki/Sample_mean_and_covariance)
* [Semivariance](/wiki/Semivariance)
* [Skewness](/wiki/Skewness)
* [Taylor's law](/wiki/Taylor's_law)
* [Weighted sample variance](/wiki/Weighted_mean#Weighted_sample_variance)

[Template:Colend](/wiki/Template:Colend)

## Notes[[edit](/index.php?title=(none)&action=edit&section=37)]

[Template:Reflist](/wiki/Template:Reflist)

[Template:Theory of probability distributions](/wiki/Template:Theory_of_probability_distributions) [Template:Statistics](/wiki/Template:Statistics)

[Template:Authority control](/wiki/Template:Authority_control)

[Category:Theory of probability distributions](/wiki/Category:Theory_of_probability_distributions) [Category:Statistical deviation and dispersion](/wiki/Category:Statistical_deviation_and_dispersion) [Category:Articles containing proofs](/wiki/Category:Articles_containing_proofs)