

Variance

From Wikipedia, the free encyclopedia

In probability theory and statistics, variance is the expectation of the squared deviation of a random variable from its 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, statistical inference, hypothesis testing, goodness of fit, Monte Carlo sampling, amongst many others. This makes it a central quantity in numerous fields such as physics, biology, chemistry, economics, and finance. The variance is the square of the standard deviation, the second central moment of a distribution, and the covariance of the random variable with itself, and it is often represented by **σ²** or **Var(***X***)**.

Contents

- 1 Definition
 - 1.1 Continuous random variable
 - 1.2 Discrete random variable
- 2 Examples
 - 2.1 Normal distribution
 - 2.2 Exponential distribution
 - 2.3 Poisson distribution
 - 2.4 Binomial distribution
 - 2.4.1 Coin toss
 - 2.5 Fair die
- 3 Properties
 - 3.1 Basic properties
 - 3.2 Sum of uncorrelated variables (Bienaymé formula)
 - 3.3 Sum of correlated variables
 - 3.4 Matrix notation for the variance of a linear combination
 - 3.5 Weighted sum of variables
 - 3.6 Product of independent variables
 - 3.7 Product of correlated variables
 - 3.8 Decomposition
 - 3.9 Formulae for the variance
 - 3.10 Calculation from the CDF
 - 3.11 Characteristic property
 - 3.12 Units of measurement
- 4 Approximating the variance of a function
- 5 Population variance and sample variance
 - 5.1 Population variance
 - 5.2 Sample variance
 - 5.3 Distribution of the sample variance
 - 5.4 Samuelson's inequality
 - 5.5 Relations with the harmonic and arithmetic means
- 6 Tests of equality of variances
- 7 History
- 8 Moment of inertia
- 9 Semivariance
- 10 Generalizations
- 11 See also
- 12 Notes

Definition

The variance of a random variable **X** is the expected value of the squared deviation from the mean of **X**, **μ** = **E[**X]**:**

Var(X**) = E[(**X** − **μ**)²].**

This definition encompasses random variables that are generated by processes that are discrete, continuous, neither, or mixed. The variance can also be thought of as the covariance of a random variable with itself:

Var(X**) = Cov(**X**, **X**).**

The variance is also equivalent to the second cumulant of a probability distribution that generates **X**. The variance is typically designated as **Var(**X**)**, **σ²**_{**X**}, or simply **σ²** (pronounced "sigma squared"). The expression for the variance can be expanded:

Var(X**) = E[(**X** − E[**X])²]
= E[**X²** − 2**X** E[**X] + (E[**X])²]
= E[**X²**] − 2E[**X] E[**X] + (E[**X])²
= E[**X²**] − (E[**X])²****************

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 if the two components of the equation are similar in magnitude. There exist numerically stable alternatives.

Continuous random variable

If the random variable **X** represents samples generated by a continuous distribution with probability density function **f(**x)****, then the population variance is given by

Var(X**) = σ² = ∫ (**x** − **μ**)² **f(**x)** ***d**x***** = ∫ **x²** **f(**x)** ***d**x***** − μ²******

where **μ** is the expected value,

μ = ∫ **x** **f(**x)** ***d**x*******

and where the integrals are definite integrals taken for **x** ranging over the range of **X**.

If a continuous distribution does not have an expected value, as is the case for the 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 whose index **k** satisfies **1** < **k** ≤ **2**.

Discrete random variable

If the generator of random variable **X** is discrete with probability mass function **x₁** ↦ **p₁**, **x₂** ↦ **p₂**, ..., **x_n** ↦ **p_n** then

Var(X**) = ∑_{i=1ⁿ **p_i · (**x_i − **μ**)²,****}**

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

$$\begin{aligned}\sigma^2 &= E(X^2) - (E(X))^2 = \frac{1}{n} \sum_{i=1}^n i^2 - \left(\frac{1}{n} \sum_{i=1}^n i\right)^2 \\ &= \frac{1}{6}(n+1)(2n+1) - \frac{1}{4}(n+1)^2 \\ &= \frac{n^2-1}{12}.\end{aligned}$$

Properties

Basic properties

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

Var(*X*) ≥ 0.

The variance of a constant random variable is zero, and if the variance of a variable in a data set is 0, then all the entries have the same value.

P(*X* = *a*) = 1 ⇔ Var(*X*) = 0.

Variance is invariant with respect to changes in a location parameter. That is, if a constant is added to all values of the variable, the variance is unchanged.

Var(*X* + *a*) = Var(*X*).

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

Var(*aX*) = *a*² Var(*X*).

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

Var(*aX* + *bY*) = *a*² Var(*X*) + *b*² Var(*Y*) + 2*ab* Cov(*X*, *Y*),

Var(*aX* − *bY*) = *a*² Var(*X*) + *b*² Var(*Y*) − 2*ab* Cov(*X*, *Y*),

where Cov(·, ·) is the covariance. In general we have for the sum of *N* random variables {*X*₁, ..., *X*_{*N*}}

$$\text{Var}\left(\sum_{i=1}^N X_i\right) = \sum_{i,j=1}^N \text{Cov}(X_i, X_j) = \sum_{i=1}^N \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j).$$

These results lead to the variance of a linear combination as:

$$\begin{aligned}\text{Var}\left(\sum_{i=1}^N a_i X_i\right) &= \sum_{i,j=1}^N a_i a_j \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^N a_i^2 \text{Var}(X_i) + \sum_{i \neq j} a_i a_j \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^N a_i^2 \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq N} a_i a_j \text{Cov}(X_i, X_j).\end{aligned}$$

If the random variables *X*₁, ..., *X*_{*N*} are such that

Cov(*X*_{*i*}, *X*_{*j*}) = 0 , ∀ (*i* ≠ *j*),

they are said to be uncorrelated. It follows immediately from the expression given earlier that if the random variables *X*₁, ..., *X*_{*N*} are uncorrelated, then the variance of their sum is equal to the sum of their variances, or, expressed symbolically:

$$\text{Var}\left(\sum_{i=1}^N X_i\right) = \sum_{i=1}^N \text{Var}(X_i).$$

Since independent random variables are always uncorrelated, the equation above holds in particular when the random variables *X*₁, ..., *X*_{*n*} 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)

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 random variables is the sum of their variances:

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i).$$

This statement is called the Bienaymé formula^[2] and was discovered in 1853. It is often made with the stronger condition that the variables are independent, 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

$$\text{Var}(\overline{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}.$$

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 of the sample mean, which is used in the central limit theorem.

Sum of correlated variables

In general, if the variables are correlated, then the variance of their sum is the sum of their covariances:

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$$

(Note: The second equality comes from the fact that Cov(*X*_{*i*},*X*_{*i*}) = Var(*X*_{*i*}).)

Here Cov(·, ·) is the 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 in classical test theory.

So if the variables have equal variance σ² and the average correlation of distinct variables is ρ, then the variance of their mean is

$$\text{Var}(\overline{X}) = \frac{\sigma^2}{n} + \frac{n-1}{n} \rho \sigma^2.$$

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. Moreover, if the variables have unit variance, for example if they are standardized, then this simplifies to

$$\mathbf{Var}(\overline{X}) = \frac{1}{n} + \frac{n-1}{n} \rho.$$

This formula is used in the 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

$$\lim_{n \rightarrow \infty} \mathbf{Var}(\overline{X}) = \rho.$$

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 states that the sample mean will converge for independent variables.

Matrix notation for the variance of a linear combination

Define ***X*** as a column vector of ***n*** random variables ***X***₁, ..., ***X***_{***n***}, and ***c*** as a column vector of ***n*** scalars ***c***₁, ..., ***c***_{***n***}. Therefore, ***c***^{*T*}***X*** is a linear combination of these random variables, where ***c***^{*T*} denotes the transpose of ***c***. Also let **Σ** be the covariance matrix of ***X***. The variance of ***c***^{*T*}***X*** is then given by:^[3]

$$\mathbf{Var}(\mathbf{c}^T \mathbf{X}) = \mathbf{c}^T \mathbf{\Sigma} \mathbf{c}.$$

Weighted sum of variables

The scaling property and the Bienaymé formula, along with the property of the covariance Cov(*aX*, *bY*) = *ab* Cov(*X*, *Y*) jointly imply that

$$\mathbf{Var}(aX \pm bY) = a^2 \mathbf{Var}(X) + b^2 \mathbf{Var}(Y) \pm 2ab \mathbf{Cov}(X, Y).$$

This implies that in a weighted sum of variables, the variable with the largest weight will have a disproportionately 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:

$$\mathbf{Var}\left(\sum_i^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \mathbf{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} a_i a_j \mathbf{Cov}(X_i, X_j)$$

Product of independent variables

If two variables *X* and *Y* are independent, the variance of their product is given by^[4]

$$\mathbf{Var}(XY) = [E(X)]^2 \mathbf{Var}(Y) + [E(Y)]^2 \mathbf{Var}(X) + \mathbf{Var}(X) \mathbf{Var}(Y).$$

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

$$\mathbf{Var}(XY) = E(X^2)E(Y^2) - [E(X)]^2[E(Y)]^2.$$

Product of correlated variables

In general, if two variables are correlated, the variance of their product is given by:

$$\begin{aligned} \mathbf{Var}(XY) &= E[X^2Y^2] - [E(XY)]^2 \\ &= \mathbf{Cov}(X^2, Y^2) + E(X^2)E(Y^2) - [E(XY)]^2 \\ &= \mathbf{Cov}(X^2, Y^2) + (\mathbf{Var}(X) + [E(X)]^2)(\mathbf{Var}(Y) + [E(Y)]^2) - [\mathbf{Cov}(X, Y) + E(X)E(Y)]^2 \end{aligned}$$

Decomposition

The general formula for variance decomposition or the law of total variance is: If ***X*** and ***Y*** are two random variables, and the variance of ***X*** exists, then

$$\mathbf{Var}[X] = \mathbf{E}_Y(\mathbf{Var}[X \mid Y]) + \mathbf{Var}_Y(\mathbf{E}[X \mid Y]).$$

where **E**(***X***|***Y***) is the conditional expectation of ***X*** given ***Y***, and **Var**(***X***|***Y***) is the conditional variance of ***X*** given ***Y***. (A more intuitive explanation is that given a particular value of ***Y***, then ***X*** follows a distribution with mean **E**(***X***|***Y***) and variance **Var**(***X***|***Y***)). As **E**(***X***|***Y***) is a function of the variable ***Y***, the outer expectation or variance is taken with respect to *Y*. The above formula tells how to find **Var**(***X***) based on the distributions of these two quantities when ***Y*** is allowed to vary. A similar formula is applied in analysis of variance, where the corresponding formula is

MS_{**total**} = ***MS***_{**between**} + ***MS***_{**within**};

here ***MS*** refers to the Mean of the Squares. In linear regression analysis the corresponding formula is

MS_{**total**} = ***MS***_{**regression**} + ***MS***_{**residual**}.

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, ***SS***):

SS_{**total**} = ***SS***_{**between**} + ***SS***_{**within**};
SS_{**total**} = ***SS***_{**regression**} + ***SS***_{**residual**}.

Formulae for the variance

A formula often used for deriving the variance of a theoretical distribution is as follows:

$$\mathbf{Var}(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2.$$

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 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.

Calculation from the CDF

The population variance for a non-negative random variable can be expressed in terms of the cumulative distribution function *F* using

$$2 \int_0^\infty u(1 - F(u)) \, du - \left(\int_0^\infty 1 - F(u) \, du\right)^2.$$

This expression can be used to calculate the variance in situations where the CDF, but not the density, can be conveniently expressed.

Characteristic property

The second moment of a random variable attains the minimum value when taken around the first moment (i.e., mean) of the random variable, i.e. **argmin**_{***m***} **E**((***X*** − ***m***)²) = **E**(***X***) . Conversely, if a continuous function **φ** satisfies **argmin**_{***m***} **E**(**φ**(***X*** − ***m***)) = **E**(***X***) for all random variables *X*, then it is necessarily of the form **φ**(***x***) = **a*****x***² + **b**, where *a* > 0. This also holds in the multidimensional case.^[5]

Units of measurement

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 or 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, is used frequently in theoretical statistics; however the expected absolute deviation tends to be more robust as it is less sensitive to outliers arising from measurement anomalies or an unduly heavy-tailed distribution.

Approximating the variance of a function

The delta method uses second-order Taylor expansions to approximate the variance of a function of one or more random variables; see Taylor expansions for the moments of functions of random variables. For example, the approximate variance of a function of one variable is given by

$$\mathbf{Var}[f(X)] \approx (f'(\mathbf{E}[X]))^2 \mathbf{Var}[X]$$

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

Population variance and sample variance

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 the mean and variance that would have been calculated from an omniscient set of observations by using an estimator equation. The estimator is a function of the sample of *n* observations drawn without observational bias from the whole 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 (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 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 for the normal distribution, and *n* − 1.5 mostly eliminates bias in 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: 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. 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 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 of the population (see 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: 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

In general, the *population variance* of a *finite* population of size *N* with values *x_i* is given by

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 = \frac{1}{N} \sum_{i=1}^N (x_i^2 - 2\mu x_i + \mu^2) = \left(\frac{1}{N} \sum_{i=1}^N x_i^2\right) - 2\mu \left(\frac{1}{N} \sum_{i=1}^N x_i\right) + \mu^2 = \left(\frac{1}{N} \sum_{i=1}^N x_i^2\right) - \mu^2$$

where the population mean is

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i.$$

Also,

$$\sigma^2 = \frac{1}{N^2} \sum_{i < j} (x_i - x_j)^2 = \frac{1}{2N^2} \sum_{i,j=1}^N (x_i - x_j)^2$$

because

$$\frac{1}{2N^2} \sum_{i,j=1}^N (x_i - x_j)^2 = \frac{1}{2N^2} \sum_{i,j=1}^N (x_i^2 - 2x_ix_j + x_j^2) = \frac{1}{2N} \sum_{j=1}^N \left(\frac{1}{N} \sum_{i=1}^N x_i^2\right) - \left(\frac{1}{N} \sum_{i=1}^N x_i^2\right) \left(\frac{1}{N} \sum_{j=1}^N x_j\right) + \frac{1}{2N} \sum_{i=1}^N \left(\frac{1}{N} \sum_{j=1}^N x_j^2\right) = \frac{1}{2} (\sigma^2 + \mu^2) - \mu^2 + \frac{1}{2} (\sigma^2 + \mu^2)$$

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

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 of the population.^[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 of *n* values *y*₁, ..., *y_n* from the population, where *n* < *N*, and estimate the variance on the basis of this sample.^[7] Directly taking the variance of the sample data gives the average of the squared deviations:

$$\sigma_{\overline{y}}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \overline{y})^2 = \left(\frac{1}{n} \sum_{i=1}^n y_i^2\right) - \overline{y}^2 = \frac{1}{n^2} \sum_{i < j} (y_i - y_j)^2.$$

Here,  ‾y denotes the sample mean:

$$\overline{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Since the *y_i* are selected randomly, both  ‾y and  σ‾y² are random variables. Their expected values can be evaluated by summing over the ensemble of all possible samples {*y_i*} from the population. For  σ‾y² this gives:

$$\begin{aligned} E[\sigma_{\overline{y}}^2] &= E\left[\frac{1}{n} \sum_{i=1}^n \left(y_i - \frac{1}{n} \sum_{j=1}^n y_j\right)^2\right] \\ &= \frac{1}{n} \sum_{i=1}^n E\left[y_i^2 - \frac{2}{n} y_i \sum_{j=1}^n y_j + \frac{1}{n^2} \sum_{j=1}^n y_j \sum_{k=1}^n y_k\right] \\ &= \frac{1}{n} \sum_{i=1}^n \left[\frac{n-2}{n} E[y_i^2] - \frac{2}{n} \sum_{j \neq i} E[y_i y_j] + \frac{1}{n^2} \sum_{j=1}^n \sum_{k \neq j} E[y_j y_k] + \frac{1}{n^2} \sum_{j=1}^n E[y_j^2]\right] \\ &= \frac{1}{n} \sum_{i=1}^n \left[\frac{n-2}{n} (\sigma^2 + \mu^2) - \frac{2}{n} (n-1) \mu^2 + \frac{1}{n^2} n(n-1) \mu^2 + \frac{1}{n} (\sigma^2 + \mu^2)\right] \\ &= \frac{n-1}{n} \sigma^2. \end{aligned}$$

Hence  σ‾y² gives an estimate of the population variance that is biased by a factor of   ‾y^{‾y}. For this reason,  σ‾y² is referred to as the *biased sample variance*. Correcting for this bias yields the *unbiased sample variance*:

$$s^2 = \frac{n}{n-1} \sigma_y^2 = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \right) = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

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, and it is also used in sample covariance and the sample standard deviation (the square root of variance). The square root is a concave function and thus introduces negative bias (by 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 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 for the function *f*(*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

Being a function of random variables, the sample variance is itself a random variable, and it is natural to study its distribution. In the case that *y*_{*i*} are independent observations from a normal distribution, Cochran's theorem shows that *s*² follows a scaled chi-squared distribution:^[8]

$$(n-1) \frac{s^2}{\sigma^2} \sim \chi^2_{n-1}.$$

As a direct consequence, it follows that

$$\mathbb{E}(s^2) = \mathbb{E}\left(\frac{\sigma^2}{n-1} \chi^2_{n-1}\right) = \sigma^2,$$

and^[9]

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

If the *y*_{*i*} are independent and identically distributed, but not necessarily normally distributed, then^{[10][11]}

$$\mathbb{E}[s^2] = \sigma^2, \quad \mathbf{Var}[s^2] = \frac{\sigma^4}{n} \left((\kappa - 1) + \frac{2}{n-1} \right) = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right),$$

where *κ* is the kurtosis of the distribution and *μ*₄ is the fourth central moment.

If the conditions of the law of large numbers hold for the squared observations, *s*² is a consistent estimator of *σ*². One can see indeed that the variance of the estimator tends asymptotically to zero. An asymptotically equivalent formula was given in Kenney and Keeping (1951:164), Rose and Smith (2002:264), and Weisstein (n.d.).^{[12][13][14]}

Samuelson's inequality

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.^[15] Values must lie within the limits ****y**** ± ****σ_y****(****n**** − 1)^{1/2}.

Relations with the harmonic and arithmetic means

It has been shown^[16] that for a sample {*y*_{*j*}} of real numbers,

$$\sigma_y^2 \leq 2 y_{\max} (A - H),$$

where *y*_{max} is the maximum of the sample, *A* is the arithmetic mean, *H* is the harmonic mean of the sample and *σ*_{*y*}² is the (biased) variance of the sample.

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

$$\sigma_y^2 \leq \frac{y_{\max} (A - H) (y_{\max} - A)}{y_{\max} - H},$$
$$\sigma_y^2 \geq \frac{y_{\min} (A - H) (A - y_{\min})}{H - y_{\min}},$$

where *y*_{min} is the minimum of the sample.^[17]

Tests of equality of variances

Testing for the equality of two or more variances is difficult. The F test and chi square tests 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, Mood test, the Klotz test and the Sukhatme test. The Sukhatme test applies to two variances and requires that both medians 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 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, the Box–Anderson test and the Moses test.

Resampling methods, which include the bootstrap and the jackknife, may be used to test the equality of variances.

History

The term *variance* was first introduced by Ronald Fisher in his 1918 paper *The Correlation Between Relatives on the Supposition of Mendelian Inheritance*.^[18]

The great body of available statistics show us that the deviations of a human measurement from its mean follow very closely the Normal Law of Errors, and, therefore, that the variability may be uniformly measured by the standard deviation corresponding to the square root of the mean square error. When there are two independent causes of variability capable of producing in an otherwise uniform population distributions with standard deviations ***σ*₁** and ***σ*₂**, it is found that the distribution, when both causes act together, has a standard deviation ***√σ*₁² + *σ*₂²**. 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...

Moment of inertia

The variance of a probability distribution is analogous to the moment of inertia in classical mechanics of a corresponding mass distribution along a line, with respect to rotation about its center of mass. It is because of this analogy that such things as the variance are called *moments* of probability distributions. The covariance matrix is related to the moment of inertia tensor for multivariate distributions. The moment of inertia of a cloud of *n* points with a covariance matrix of ***Σ*** is given by

I = ***n***(***1*_{3×3}** tr(***Σ***) − ***Σ***).

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

Σ =

[

10

0

0

0

0.1

0

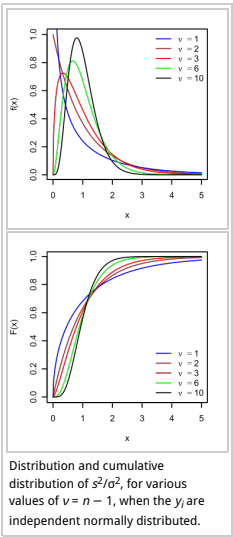
0

0

0.1

]

.



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

$$I = \boldsymbol{\mathfrak{n}} \begin{bmatrix} \mathbf{0.2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{10.1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{10.1} \end{bmatrix}.$$

Semivariance

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 in an investments context. For skewed distributions, the semivariance can provide additional information that a variance does not.

Generalizations

If ***x*** is a scalar complex-valued random variable, with values in **C**, then its variance is **E**((***x*** − ***μ***)(***x*** − ***μ***^{*})), where ***x***^{*} is the complex conjugate of ***x***. This variance is a real scalar.

If ***X*** is a vector-valued random variable, with values in **R**^{*n*}, and thought of as a column vector, then the natural generalization of variance is **E**((***X*** − ***μ***)(***X*** − ***μ***^T)), where ***μ*** = **E**(***X***) and ***X***^T is the transpose of ***X***, and so is a row vector. The result is a positive semi-definite square matrix, commonly referred to as the variance-covariance matrix.

If ***X*** is a vector- and complex-valued random variable, with values in **C**^{*n*}, then the generalization of its variance is **E**((***X*** − ***μ***)(***X*** − ***μ***[†])), where ***X***[†] is the conjugate transpose of ***X***. This matrix is also positive semi-definite and square.

See also

- Algorithms for calculating variance
 - Average absolute deviation
 - Bhatia–Davis inequality
 - Common-method variance
 - Correlation
 - Covariance
 - Chebyshev's inequality
 - Distance variance
- Estimation of covariance matrices
 - Explained variance
 - Homoscedasticity
 - Mean absolute error
 - Mean difference
 - Mean preserving spread
 - Pooled variance (also known as combined, composite, or overall variance)
- Popoviciu's inequality on variances
 - Qualitative variation
 - Sample mean and covariance
 - Semivariance
 - Skewness
 - Taylor's law
 - Weighted sample variance

Notes

1. Yuli Zhang,Huaiyu Wu,Lei Cheng (June 2012). *Some new deformation formulas about variance and covariance*. Proceedings of 4th International Conference on Modelling, Identification and Control(ICMIC2012). pp. 987–992.

2. Loeve, M. (1977) "Probability Theory", *Graduate Texts in Mathematics*, Volume 45, 4th edition, Springer-Verlag, p. 12.

3. Johnson, Richard; Wichern, Dean (2001). *Applied Multivariate Statistical Analysis*. Prentice Hall. p. 76. ISBN 0-13-187715-1

4. Goodman, Leo A., "On the exact variance of products," *Journal of the American Statistical Association*, December 1960, 708–713. DOI: 10.2307/2281592 (http://dx.doi.org/DOI:10.2307/2281592)

5. Kagan, A.; Shepp, L. A. (1998). "Why the variance?". *Statistics & Probability Letters*. 38 (4): 329–333. doi:10.1016/S0167-7152(98)00041-8.

6. Navidi, William (2006) *Statistics for Engineers and Scientists*, McGraw-Hill, pg 14.

7. Montgomery, D. C. and Runger, G. C. (1994) *Applied statistics and probability for engineers*, page 201. John Wiley & Sons New York

8. Knight K. (2000), *Mathematical Statistics*, Chapman and Hall, New York. (proposition 2.11)

9. Casella and Berger (2002) *Statistical Inference*, Example 7.3.3, p. 331

10. Cho, Eungchun; Cho, Moon Jung; Eltinge, John (2005) The Variance of Sample Variance From a Finite Population. International Journal of Pure and Applied Mathematics 21 (3): 387-394. http://www.ijpam.eu/contents/2005-21-3/10/10.pdf

11. Cho, Eungchun; Cho, Moon Jung (2009) Variance of Sample Variance With Replacement. International Journal of Pure and Applied Mathematics 52 (1): 43-47. http://www.ijpam.eu/contents/2009-52-1/5/5.pdf

12. Kenney, John F.; Keeping, E.S. (1951) Mathematics of Statistics. Part Two. 2nd ed. D. Van Nostrand Company, Inc. Princeton: New Jersey. http://krishikosh.egranth.ac.in/bitstream/1/2025521/1/G2257.pdf

13. Rose, Colin; Smith, Murray D. (2002) Mathematical Statistics with Mathematica. Springer-Verlag, New York. http://www.mathstatica.com/book/Mathematical_Statistics_with_Mathematica.pdf

14. Weisstein, Eric W. (n.d.) Sample Variance Distribution. MathWorld—A Wolfram Web Resource. http://mathworld.wolfram.com/SampleVarianceDistribution.html

15. Samuelson, Paul (1968). "How Deviant Can You Be?". *Journal of the American Statistical Association*. 63 (324): 1522–1525. doi:10.1080/01621459.1968.10480944. JSTOR 2285901.

16. Mercer, A. McD. (2000). "Bounds for A–G, A–H, G–H, and a family of inequalities of Ky Fan's type, using a general method". *J. Math. Anal. Appl.* 243 (1): 163–173. doi:10.1006/jmaa.1999.6688.

17. Sharma, R. (2008). "Some more inequalities for arithmetic mean, harmonic mean and variance". *J. Math. Inequalities*. 2 (1): 109–114. doi:10.7153/jmi-02-11.

18. Ronald Fisher (1918) The correlation between relatives on the supposition of Mendelian Inheritance (http://digital.library.adelaide.edu.au/dspace/bitstream/2440/15097/1/9.pdf)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Variance&oldid=738238933"

Categories: Theory of probability distributions | Statistical deviation and dispersion

- This page was last modified on 7 September 2016, at 19:18.
 - Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

