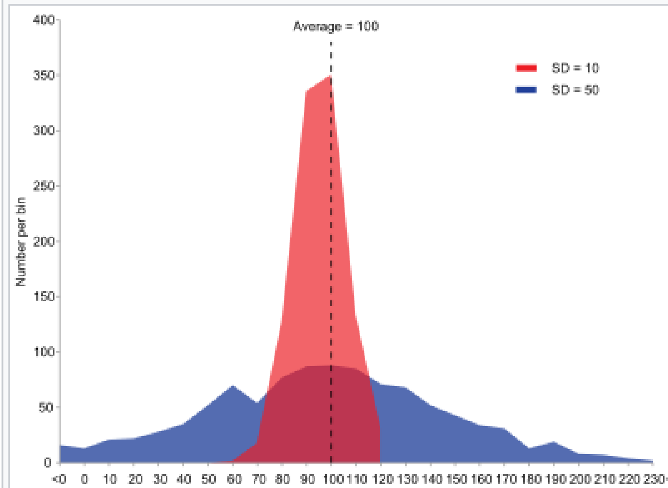
Variance

From Wikipedia, the free encyclopedia

This article is about the mathematical concept. For other uses, see Variance (disambiguation).

In probability theory and statistics, **variance** is the expectation of the squared deviation of a random variable from its mean. Informally, it measures how far a set of numbers is spread out from their average value. Variance has a central role in statistics, where some ideas that use it include descriptive statistics, statistical inference, hypothesis testing, goodness of fit, and Monte Carlo sampling. Variance is an important tool in the sciences, where statistical analysis of data is common. 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 σ^2, s^2, or $\operatorname{Var}(X)$.



Example of samples from two populations with the same mean but different variances. The red population has mean 100 and variance 100 (SD=10) while the blue population has mean 100 and variance 2500 (SD=50).

Contents [hide]

Definition [edit]

The variance of a random variable X is the expected value of the squared deviation from the mean of X, $\mu = \operatorname{E}[X]$:

$$\operatorname{Var}(X) = \operatorname{E}\left[(X-\mu)^2\right].$$

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:

$$\operatorname{Var}(X) = \operatorname{Cov}(X, X).$$

The variance is also equivalent to the second cumulant of a probability distribution that generates X. The variance is typically designated as $\operatorname{Var}(X)$, σ_X^2, or simply σ^2 (pronounced "sigma squared"). The expression for the variance can be expanded:

$$\begin{aligned}\operatorname{Var}(X) &= \operatorname{E}\left[(X-\operatorname{E}[X])^2\right] \\ &= \operatorname{E}\left[X^2 - 2X\operatorname{E}[X] + \operatorname{E}[X]^2\right] \\ &= \operatorname{E}\left[X^2\right] - 2\operatorname{E}[X]\operatorname{E}[X] + \operatorname{E}[X]^2 \\ &= \operatorname{E}\left[X^2\right] - \operatorname{E}[X]^2\end{aligned}$$

In other words, the variance of X is equal to the mean of the square of X minus the square of the mean of X. This equation should not be used for computations using floating point arithmetic because it suffers from catastrophic cancellation if the two components of the equation are similar in magnitude. There exist numerically stable alternatives.

Discrete random variable [edit]

If the generator of random variable X is discrete with probability mass function $x_1 \mapsto p_1, x_2 \mapsto p_2, \ldots, x_n \mapsto p_n$ then

$$\operatorname{Var}(X) = \sum_{i=1}^{n} p_i \cdot (x_i - \mu)^2,$$

or equivalently

$$\operatorname{Var}(X) = \left(\sum_{i=1}^{n} p_i x_i^2\right) - \mu^2,$$

where μ is the expected value, i.e.,

$$\mu = \sum_{i=1}^{n} p_i x_i.$$

(When such a discrete 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 n equally likely values can be written as

$$\operatorname{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2,$$

where μ is the average value, i.e.,

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

The variance of a set of n 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]

$$\operatorname{Var}(X) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{2}(x_i - x_j)^2 = \frac{1}{n^2}\sum_{i}\sum_{j>i}(x_i - x_j)^2.$$

Absolutely continuous random variable [edit]

If the random variable X has a probability density function $f(x)$, and $F(x)$ is the corresponding cumulative distribution function, then

$$\begin{aligned}
\operatorname{Var}(X) = \sigma^2 &= \int_{\mathbb{R}} (x-\mu)^2 f(x)\,dx \\
&= \int_{\mathbb{R}} x^2 f(x)\,dx - 2\mu\int_{\mathbb{R}} x f(x)\,dx + \mu^2\int_{\mathbb{R}} f(x)\,dx \\
&= \int_{\mathbb{R}} x^2\,dF(x) - 2\mu\int_{\mathbb{R}} x\,dF(x) + \mu^2\int_{\mathbb{R}}\,dF(x) \\
&= \int_{\mathbb{R}} x^2\,dF(x) - 2\mu\cdot\mu + \mu^2\cdot 1 \\
&= \int_{\mathbb{R}} x^2\,dF(x) - \mu^2,
\end{aligned}$$

or equivalently,

$$\operatorname{Var}(X) = \int_{\mathbb{R}} x^2 f(x)\,dx - \mu^2,$$

where μ is the expected value of X given by

$$\mu = \int_{\mathbb{R}} x f(x)\,dx = \int_{\mathbb{R}} x\,dF(x).$$

In these formulas, the integrals with respect to dx and $dF(x)$ are Lebesgue and Lebesgue–Stieltjes integrals, respectively.

If the function $x^2 f(x)$ is Riemann-integrable on every finite interval $[a,b] \subset \mathbb{R}$, then

$$\operatorname{Var}(X) = \int_{-\infty}^{+\infty} x^2 f(x)\,dx - \mu^2,$$

where the integral is an improper Riemann integral.

Examples [edit]

Exponential distribution [edit]

The exponential distribution with parameter λ is a continuous distribution whose probability density function is given by

$$f(x) = \lambda e^{-\lambda x}$$

on the interval $[0, \infty)$. Its mean can be shown to be

$$\mathrm{E}[X] = \int_0^\infty \lambda x e^{-\lambda x}\, dx = \frac{1}{\lambda}.$$

Using integration by parts and making use of the expected value already calculated:

$$\begin{aligned}\mathrm{E}\left[X^2\right] &= \int_0^\infty \lambda x^2 e^{-\lambda x}\, dx \\ &= \left[-x^2 e^{-\lambda x}\right]_0^\infty + \int_0^\infty 2x e^{-\lambda x}\, dx \\ &= 0 + \frac{2}{\lambda}\mathrm{E}[X] \\ &= \frac{2}{\lambda^2}.\end{aligned}$$

Thus, the variance of X is given by

$$\operatorname{Var}(X) = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

Fair die [edit]

A fair six-sided die can be modeled as a discrete random variable, X, with outcomes 1 through 6, each with equal probability 1/6. The expected value of X is $(1+2+3+4+5+6)/6 = 7/2.$ Therefore, the variance of X is

$$\begin{aligned}\operatorname{Var}(X) &= \sum_{i=1}^{6} \frac{1}{6}\left(i - \frac{7}{2}\right)^2 \\ &= \frac{1}{6}\left((-5/2)^2 + (-3/2)^2 + (-1/2)^2 + (1/2)^2 + (3/2)^2 + (5/2)^2\right) \\ &= \frac{35}{12} \approx 2.92.\end{aligned}$$

The general formula for the variance of the outcome, X, of an n-sided die is

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

Commonly used probability distributions [edit]

The following table lists the variance for some commonly used probability distributions.

Name of the probability distribution	Probability distribution function	Mean	Variance
Binomial distribution	$\Pr\,(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Geometric distribution	$\Pr\,(X=k) = (1-p)^{k-1} p$	$\frac{1}{p}$	$\frac{(1-p)}{p^2}$
Normal distribution	$f\left(x \mid \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Uniform distribution (continuous)	$f(x \mid a, b) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential distribution	$f(x \mid \lambda) = \lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Properties [edit]

Basic properties [edit]

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

$$\operatorname{Var}(X) \ge 0.$$

The variance of a constant is zero.

$$\operatorname{Var}(a) = 0.$$

If the variance of a random variable is 0, then it is a constant. That is, it always has the same value:

$$\operatorname{Var}(X) = 0 \iff \exists a : P(X = a) = 1.$$

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:

$$\operatorname{Var}(X + a) = \operatorname{Var}(X).$$

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

$$\operatorname{Var}(aX) = a^2 \operatorname{Var}(X).$$

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

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

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

where $\operatorname{Cov}(X, Y)$ is the covariance. In general we have for the sum of N random variables $\{X_1, \dots, X_N\}$:

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

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

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

If the random variables $X_1, \dots, X_N$ are such that

$$\operatorname{Cov}(X_i, X_j) = 0 \ , \ \forall \ (i \neq j),$$

they are said to be uncorrelated. It follows immediately from the expression given earlier that if the random variables $X_1, \dots, X_N$ are uncorrelated, then the variance of their sum is equal to the sum of their variances, or, expressed symbolically:

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

Since independent random variables are always uncorrelated, the equation above holds in particular when the random variables $X_1, \dots, X_n$ are independent. Thus independence is sufficient but not necessary for the variance of the sum to equal the sum of the variances.

Issues of finiteness [edit]

If a distribution does not have a finite expected value, as is the case for the Cauchy distribution, then the variance cannot be finite either. However, some distributions may not have a finite variance despite their expected value being finite. An example is a Pareto distribution whose index k satisfies $1 < k \leq 2.$

Sum of uncorrelated variables (Bienaymé formula) [edit]

See also: Sum of normally distributed random variables

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:

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

This statement is called the Bienaymé formula[2] and was discovered in 1853.[3][4] 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 σ^2, then, since division by *n* is a linear transformation, this formula immediately implies that the variance of their mean is

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

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.

To prove the initial statement, it suffices to show that

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y).$$

The general result then follows by induction. Starting with the definition,

$$\begin{aligned}\operatorname{Var}(X+Y) &= \operatorname{E}\left[(X+Y)^2\right] - (\operatorname{E}[X+Y])^2\\ &= \operatorname{E}\left[X^2 + 2XY + Y^2\right] - (\operatorname{E}[X] + \operatorname{E}[Y])^2.\end{aligned}$$

Using the linearity of the expectation operator and the assumption of independence (or uncorrelatedness) of X and Y, this further simplifies as follows:

$$\begin{aligned}\operatorname{Var}(X+Y) &= \operatorname{E}\left[X^2\right] + 2\operatorname{E}[XY] + \operatorname{E}\left[Y^2\right] - \left(\operatorname{E}[X]^2 + 2\operatorname{E}[X]\operatorname{E}[Y] + \operatorname{E}[Y]^2\right) \\ &= \operatorname{E}\left[X^2\right] + \operatorname{E}\left[Y^2\right] - \operatorname{E}[X]^2 - \operatorname{E}[Y]^2 \\ &= \operatorname{Var}(X) + \operatorname{Var}(Y).\end{aligned}$$

Sum of correlated variables [edit]

With correlation and fixed sample size [edit]

In general the variance of the sum of n variables is the sum of their covariances:

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

(Note: The second equality comes from the fact that $\operatorname{Cov}(X_i,X_i) = \operatorname{Var}(X_i)$.)

Here $\operatorname{Cov}(\cdot,\cdot)$ 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. The next expression states equivalently that the variance of the sum is the sum of the diagonal of covariance matrix plus two times the sum of its upper triangular elements (or its lower triangular elements); this emphasizes that the covariance matrix is symmetric. This formula is used in the theory of Cronbach's alpha in 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

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

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

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

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

$$\lim_{n\to\infty} \operatorname{Var}\left(\overline{X}\right) = \rho.$$

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.

I.i.d. with random sample size [edit]

There are cases when a sample is taken without knowing, in advance, how many observations will be acceptable according to some criterion. In such cases, the sample size $\boldsymbol{N}$ is a random variable whose variation adds to the variation of $\boldsymbol{X}$, such that,

$\operatorname{Var}(\sum \boldsymbol{X}) = \operatorname{E}(\boldsymbol{N})\operatorname{Var}(\boldsymbol{X}) + \operatorname{Var}(\boldsymbol{N})\operatorname{E}^2(\boldsymbol{X})$.[5]

If $\boldsymbol{N}$ has a Poisson distribution, then $\operatorname{E}(\boldsymbol{N}) = \operatorname{Var}(\boldsymbol{N})$ with estimator $\boldsymbol{N} = n$. So, the estimator of $\operatorname{Var}(\sum \boldsymbol{X})$ becomes $nS^2_{\boldsymbol{X}} + n\overline{\boldsymbol{X}}^2$ giving

standard error$(\overline{\boldsymbol{X}}) = \sqrt{[(S^2_{\boldsymbol{X}} + \overline{\boldsymbol{X}}^2)/n]}$.

Matrix notation for the variance of a linear combination [edit]

Define X as a column vector of n random variables $X_1, \ldots, X_n$, and c as a column vector of n scalars $c_1, \ldots, c_n$. Therefore, $c^\mathsf{T} X$ 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 $c^\mathsf{T} X$ is then given by:[6]

$$\operatorname{Var}\left(c^\mathsf{T} X\right) = c^\mathsf{T} \Sigma c.$$

This implies that the variance of the mean can be written as (with a column vector of ones)

$$\operatorname{Var}(\bar{x}) = \operatorname{Var}\left(\frac{1}{n} 1' X\right) = \frac{1}{n^2} 1' \Sigma 1.$$

Weighted sum of variables [edit]

Not to be confused with Weighted variance.

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

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

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:

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

Product of independent variables [edit]

If two variables X and Y are independent, the variance of their product is given by[7]

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

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

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

Product of statistically dependent variables [edit]

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

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

Decomposition [edit]

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

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

The conditional expectation $\operatorname{E}(X \mid Y)$ of X given Y, and the conditional variance $\operatorname{Var}(X \mid Y)$ may be understood as follows. Given any particular value *y* of the random variable *Y*, there is a conditional expectation $\operatorname{E}(X \mid Y = y)$ given the event *Y* = *y*. This quantity depends on the particular value *y*; it is a function $g(y) = \operatorname{E}(X \mid Y = y)$. That same function evaluated at the random variable *Y* is the conditional expectation $\operatorname{E}(X \mid Y) = g(Y)$.

In particular, if Y is a discrete random variable assuming possible values $y_1, y_2, y_3 \ldots$ with corresponding probabilities $p_1, p_2, p_3 \ldots,$, then in the formula for total variance, the first term on the right-hand side becomes

$$\mathrm{E}(\operatorname{Var}[X \mid Y]) = \sum_i p_i \sigma_i^2,$$

where $\sigma_i^2 = \operatorname{Var}[X \mid Y = y_i]$. Similarly, the second term on the right-hand side becomes

$$\operatorname{Var}(\mathrm{E}[X \mid Y]) = \sum_i p_i \mu_i^2 - \left(\sum_i p_i \mu_i\right)^2 = \sum_i p_i \mu_i^2 - \mu^2,$$

where $\mu_i = \mathrm{E}[X \mid Y = y_i]$ and $\mu = \sum_i p_i \mu_i$. Thus the total variance is given by

$$\operatorname{Var}[X] = \sum_i p_i \sigma_i^2 + \left(\sum_i p_i \mu_i^2 - \mu^2\right).$$

A similar formula is applied in analysis of variance, where the corresponding formula is

$$MS_{\text{total}} = MS_{\text{between}} + MS_{\text{within}};$$

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

$$MS_{\text{total}} = MS_{\text{regression}} + MS_{\text{residual}}.$$

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):

$$SS_{\text{total}} = SS_{\text{between}} + SS_{\text{within}},$$
$$SS_{\text{total}} = SS_{\text{regression}} + SS_{\text{residual}}.$$

Calculation from the CDF [edit]

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 [edit]

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. $\mathrm{argmin}_m \, \mathrm{E}\left(\left(X - m\right)^2\right) = \mathrm{E}(X)$. Conversely, if a continuous function φ satisfies $\mathrm{argmin}_m \, \mathrm{E}(\varphi(X - m)) = \mathrm{E}(X)$ for all random variables *X*, then it is necessarily of the form $\varphi(x) = ax^2 + b$, where *a* > 0. This also holds in the multidimensional case.[8]

Units of measurement [edit]

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 $\sqrt{2.9} \approx 1.7$, 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 [edit]

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

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

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

Population variance and sample variance [edit]

See also: Unbiased estimation of standard deviation

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 [edit]

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

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

where the population mean is

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

The population variance can also be computed using

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

This is true because

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

The population variance 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]

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.[9] 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.[10] Directly taking the variance of the sample data gives the average of the squared deviations:

$$\sigma_Y^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2 = \left(\frac{1}{n}\sum_{i=1}^{n}Y_i^2\right)-\overline{Y}^2 = \frac{1}{n^2}\sum_{i,j\,:\,i<j}(Y_i-Y_j)^2.$$

Here, $\overline{Y}$ denotes the sample mean:

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n}Y_i.$$

Since the Y_i are selected randomly, both $\overline{Y}$ and σ_Y^2 are random variables. Their expected values can be evaluated by averaging over the ensemble of all possible samples $\{Y_i\}$ of size n from the population. For σ_Y^2 this gives:

$$\begin{aligned}
\mathrm{E}[\sigma_Y^2] &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \frac{1}{n}\sum_{j=1}^{n}Y_j\right)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[Y_i^2 - \frac{2}{n}Y_i\sum_{j=1}^{n}Y_j + \frac{1}{n^2}\sum_{j=1}^{n}Y_j\sum_{k=1}^{n}Y_k\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{n-2}{n}\mathrm{E}\left[Y_i^2\right] - \frac{2}{n}\sum_{j\neq i}\mathrm{E}[Y_iY_j] + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k\neq j}^{n}\mathrm{E}[Y_jY_k] + \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{E}\left[Y_j^2\right]\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{n-2}{n}\left(\sigma^2+\mu^2\right) - \frac{2}{n}(n-1)\mu^2 + \frac{1}{n^2}n(n-1)\mu^2 + \frac{1}{n}\left(\sigma^2+\mu^2\right)\right] \\
&= \frac{n-1}{n}\sigma^2.
\end{aligned}$$

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

$$s^2 = \frac{n}{n-1}\sigma_Y^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2\right) = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$$

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 [edit]

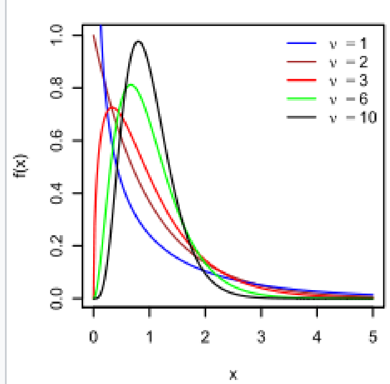
Being a function of random variables, the sample variance is itself a random variable, and it is natural to study its distribution. In the case that Y_i are independent observations from a normal distribution, Cochran's theorem shows that s^2 follows a scaled chi-squared distribution:[11]

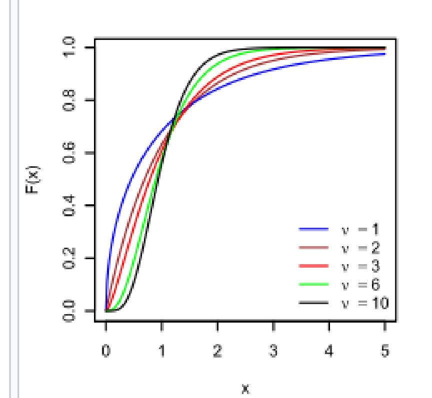
$$(n-1)\frac{s^2}{\sigma^2} \sim \chi^2_{n-1}.$$

As a direct consequence, it follows that

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

and[12]





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

If the Y_i are independent and identically distributed, but not necessarily normally distributed, then[13]

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

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.).[14][15][16]

Samuelson's inequality [edit]

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.[17] Values must lie within the limits $\bar{y} \pm \sigma_Y (n-1)^{1/2}$.

Relations with the harmonic and arithmetic means [edit]

It has been shown[18] that for a sample $\{y_i\}$ of positive real numbers,

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

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

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

where $y_{\min}$ is the minimum of the sample.[19]

Tests of equality of variances [edit]

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 [edit]

The term *variance* was first introduced by Ronald Fisher in his 1918 paper *The Correlation Between Relatives on the Supposition of Mendelian Inheritance*:[20]

> 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 $\sqrt{\sigma_1^2+\sigma_2^2}$. 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 [edit]

See also: Moment (physics) § Examples

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.[*citation needed*] It is because of this analogy that such things as the variance are called *moments* of probability distributions.[*citation needed*] 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[*citation needed*]

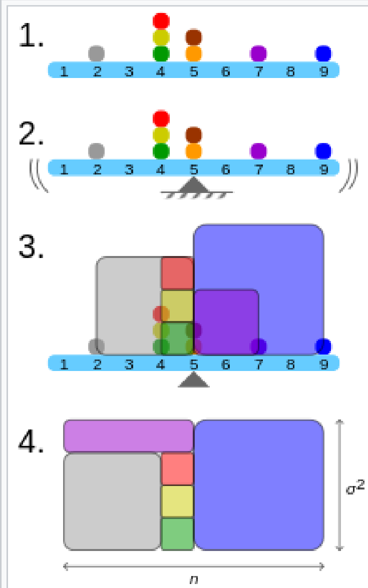
$$I=n\left(\mathbf{1}_{3\times 3}\operatorname{tr}(\Sigma)-\Sigma\right).$$

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

$$\Sigma=\begin{bmatrix}10&0&0\\0&0.1&0\\0&0&0.1\end{bmatrix}.$$

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

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



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.

Semivariance [edit]

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:

$$\text{Semivariance} = \frac{1}{n} \sum_{i:x_i<\mu} (x_i - \mu)^2$$

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.[21]

For inequalities associated with the semivariance, see Chebyshev's inequality § Semivariances.

Generalizations [edit]

For complex variables [edit]

If x is a scalar complex-valued random variable, with values in $\mathbb{C}$, then its variance is $\operatorname{E}\left[(x-\mu)(x-\mu)^*\right]$, where x^* is the complex conjugate of x. This variance is a real scalar.

For vector-valued random variables [edit]

As a matrix [edit]

If X is a vector-valued random variable, with values in $\mathbb{R}^n$, and thought of as a column vector, then a natural generalization of variance is $\operatorname{E}\left[(X-\mu)(X-\mu)^{\operatorname{T}}\right]$, where $\mu = \operatorname{E}(X)$ 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 (or simply as the *covariance matrix*).

If X is a vector- and complex-valued random variable, with values in $\mathbb{C}^n$, then the covariance matrix is $\operatorname{E}\left[(X-\mu)(X-\mu)^\dagger\right]$, where $X^\dagger$ is the conjugate transpose of X.[*citation needed*] This matrix is also positive semi-definite and square.

As a scalar [edit]

Another generalization of variance for vector-valued random variables X, which results in a scalar value rather than in a matrix, is the generalized variance $\det(C)$, the determinant of the covariance matrix. The generalized variance can be shown to be related to the multidimensional scatter of points around their mean.[22]

A different generalization is obtained by considering the Euclidean distance between the random variable and its mean. This results in $\operatorname{E}\left[(X-\mu)^{\operatorname{T}}(X-\mu)\right] = \operatorname{tr}(C)$, which is the trace of the covariance matrix.

See also [edit]

- Measures for Statistical Dispersion
- Correlation
- Distance variance
- Explained variance
- Pooled variance
- Homoscedasticity
- Bhatia–Davis inequality
- Popoviciu's inequality on variances

Mathematics portal

Look up ***variance*** in Wiktionary, the free dictionary.

References [edit]

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

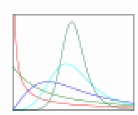
3. ^ Bienaymé, I.-J. (1853) "Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés", *Comptes rendus de l'Académie des sciences Paris*, 37, p. 309–317; digital copy available [1]
4. ^ Bienaymé, I.-J. (1867) "Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés", *Journal de Mathématiques Pures et Appliquées, Série 2*, Tome 12, p. 158–167; digital copy available [2][3]
5. ^ Cornell, J R, and Benjamin, C A, *Probability, Statistics, and Decisions for Civil Engineers,* McGraw-Hill, NY, 1970, pp.178-9.
6. ^ Johnson, Richard; Wichern, Dean (2001). *Applied Multivariate Statistical Analysis*. Prentice Hall. p. 76. ISBN 0-13-187715-1.
7. ^ Goodman, Leo A. (December 1960). "On the Exact Variance of Products". *Journal of the American Statistical Association*. **55** (292): 708–713. doi:10.2307/2281592. JSTOR 2281592.
8. ^ Kagan, A.; Shepp, L. A. (1998). "Why the variance?". *Statistics & Probability Letters*. **38** (4): 329–333. doi:10.1016/S0167-7152(98)00041-8.
9. ^ Navidi, William (2006) *Statistics for Engineers and Scientists*, McGraw-Hill, pg 14.
10. ^ Montgomery, D. C. and Runger, G. C. (1994) *Applied statistics and probability for engineers*, page 201. John Wiley & Sons New York
11. ^ Knight K. (2000), *Mathematical Statistics*, Chapman and Hall, New York. (proposition 2.11)
12. ^ Casella and Berger (2002) *Statistical Inference*, Example 7.3.3, p. 331[*full citation needed*]
13. ^ 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
14. ^ 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
15. ^ Rose, Colin; Smith, Murray D. (2002) Mathematical Statistics with Mathematica. Springer-Verlag, New York. http://www.mathstatica.com/book/Mathematical_Statistics_with_Mathematica.pdf
16. ^ Weisstein, Eric W. (n.d.) Sample Variance Distribution. MathWorld—A Wolfram Web Resource. http://mathworld.wolfram.com/SampleVarianceDistribution.html
17. ^ 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.
18. ^ 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.
19. ^ Sharma, R. (2008). "Some more inequalities for arithmetic mean, harmonic mean and variance". *Journal of Mathematical Inequalities*. **2** (1): 109–114. CiteSeerX 10.1.1.551.9397. doi:10.7153/jmi-02-11.
20. ^ Ronald Fisher (1918) The correlation between relatives on the supposition of Mendelian Inheritance
21. ^ Fama, Eugene F.; French, Kenneth R. (2010-04-21). "Q&A: Semi-Variance: A Better Risk Measure?". *Fama/French Forum*.
22. ^ Kocherlakota, S.; Kocherlakota, K. (2004). "Generalized Variance". *Encyclopedia of Statistical Sciences*. Wiley Online Library. doi:10.1002/0471667196.ess0869. ISBN 0471667196.

V · T · E **Theory of probability distributions** [hide]

probability mass function (pmf) · probability density function (pdf) · cumulative distribution function (cdf) · quantile function

raw moment · central moment · mean · **variance** · standard deviation · skewness · kurtosis · L-moment

moment-generating function (mgf) · characteristic function · probability-generating function (pgf) · cumulant · combinant



V · T · E **Statistics** [show]

Authority control GND: 4078739-4 · NDL: 00561029

Categories: Moment (mathematics) | Statistical deviation and dispersion