**Mean squared error**

From Wikipedia, the free encyclopedia

"Mean squared deviation" redirects here. It is not to be confused with [Mean squared displacement](https://en.wikipedia.org/wiki/Mean_squared_displacement).

In [statistics](https://en.wikipedia.org/wiki/Statistics), the **mean squared error** (**MSE**) or **mean squared deviation** (**MSD**) of an [estimator](https://en.wikipedia.org/wiki/Estimator) measures the [average](https://en.wikipedia.org/wiki/Expected_value) of the squares of the [errors](https://en.wikipedia.org/wiki/Error_%28statistics%29) or [deviations](https://en.wikipedia.org/wiki/Deviation_%28statistics%29), that is, the difference between the estimator and what is estimated. MSE is a [risk function](https://en.wikipedia.org/wiki/Risk_function), corresponding to the [expected value](https://en.wikipedia.org/wiki/Expected_value) of the squared error loss or quadratic loss. The difference occurs because of [randomness](https://en.wikipedia.org/wiki/Randomness) or because the estimator [doesn't account for information](https://en.wikipedia.org/wiki/Omitted-variable_bias) that could produce a more accurate estimate.[[1]](https://en.wikipedia.org/wiki/Mean_squared_error#cite_note-pointEstimation-1)

The MSE is the second [moment](https://en.wikipedia.org/wiki/Moment_%28mathematics%29) (about the origin) of the error, and thus incorporates both the [variance](https://en.wikipedia.org/wiki/Variance) of the estimator and its [bias](https://en.wikipedia.org/wiki/Bias_of_an_estimator). For an [unbiased estimator](https://en.wikipedia.org/wiki/Unbiased_estimator), the MSE is the variance of the estimator. Like the variance, MSE has the same units of measurement as the square of the quantity being estimated. In an analogy to [standard deviation](https://en.wikipedia.org/wiki/Standard_deviation), taking the square root of MSE yields the root-mean-square error or [root-mean-square deviation](https://en.wikipedia.org/wiki/Root-mean-square_deviation) (RMSE or RMSD), which has the same units as the quantity being estimated; for an unbiased estimator, the RMSE is the square root of the [variance](https://en.wikipedia.org/wiki/Variance), known as the [standard deviation](https://en.wikipedia.org/wiki/Standard_deviation).

**Contents**

* [1 Definition and basic properties](https://en.wikipedia.org/wiki/Mean_squared_error#Definition_and_basic_properties)
  + [1.1 Predictor](https://en.wikipedia.org/wiki/Mean_squared_error#Predictor)
  + [1.2 Estimator](https://en.wikipedia.org/wiki/Mean_squared_error#Estimator)
    - [1.2.1 Proof of variance and bias relationship](https://en.wikipedia.org/wiki/Mean_squared_error#Proof_of_variance_and_bias_relationship)
* [2 Regression](https://en.wikipedia.org/wiki/Mean_squared_error#Regression)
* [3 Examples](https://en.wikipedia.org/wiki/Mean_squared_error#Examples)
  + [3.1 Mean](https://en.wikipedia.org/wiki/Mean_squared_error#Mean)
  + [3.2 Variance](https://en.wikipedia.org/wiki/Mean_squared_error#Variance)
  + [3.3 Gaussian distribution](https://en.wikipedia.org/wiki/Mean_squared_error#Gaussian_distribution)
* [4 Interpretation](https://en.wikipedia.org/wiki/Mean_squared_error#Interpretation)
* [5 Applications](https://en.wikipedia.org/wiki/Mean_squared_error#Applications)
* [6 Loss function](https://en.wikipedia.org/wiki/Mean_squared_error#Loss_function)
  + [6.1 Criticism](https://en.wikipedia.org/wiki/Mean_squared_error#Criticism)
* [7 See also](https://en.wikipedia.org/wiki/Mean_squared_error#See_also)
* [8 Notes](https://en.wikipedia.org/wiki/Mean_squared_error#Notes)
* [9 References](https://en.wikipedia.org/wiki/Mean_squared_error#References)

**Definition and basic properties**

The MSE assesses the quality of an [**estimator**](https://en.wikipedia.org/wiki/Estimator) (i.e., a [mathematical function](https://en.wikipedia.org/wiki/Mathematical_function) mapping a [sample](https://en.wikipedia.org/wiki/Sample_%28statistics%29) of data to a [parameter](https://en.wikipedia.org/wiki/Statistical_parameter) of the [population](https://en.wikipedia.org/wiki/Statistical_population) from which the data is sampled) or a **predictor** (i.e., a function mapping arbitrary inputs to a sample of values of some [random variable](https://en.wikipedia.org/wiki/Random_variable)). Definition of an MSE differs according to whether one is describing an estimator or a predictor.

**Predictor**

If Y ^ {\displaystyle {\hat {Y}}} is a vector of n {\displaystyle n} predictions, and Y {\displaystyle Y} is the vector of observed values corresponding to the inputs to the function which generated the predictions, then the MSE of the predictor can be estimated by

MSE = 1 n ∑ i = 1 n ( Y i ^ − Y i ) 2 {\displaystyle \operatorname {MSE} ={\frac {1}{n}}\sum \_{i=1}^{n}({\hat {Y\_{i}}}-Y\_{i})^{2}}

I.e., the MSE is the *mean* ( 1 n ∑ i = 1 n {\displaystyle {\frac {1}{n}}\sum \_{i=1}^{n}} ) of the *square of the errors* ( ( Y i ^ − Y i ) 2 {\displaystyle ({\hat {Y\_{i}}}-Y\_{i})^{2}} ). This is an easily computable quantity for a particular sample (and hence is sample-dependent).

**Estimator**

The MSE of an estimator θ ^ {\displaystyle {\hat {\theta }}} with respect to an unknown parameter θ {\displaystyle \theta } is defined as

MSE ⁡ ( θ ^ ) = E ⁡ [ ( θ ^ − θ ) 2 ] . {\displaystyle \operatorname {MSE} ({\hat {\theta }})=\operatorname {E} {\big [}({\hat {\theta }}-\theta )^{2}{\big ]}.}

This definition depends on the unknown parameter, and the MSE in this sense is a property of an estimator. Since an MSE is an [expectation](https://en.wikipedia.org/wiki/Expected_value), it is not technically a random variable. That being said, the MSE could be a function of unknown parameters, in which case any *estimator* of the MSE based on estimates of these parameters would be a function of the data and thus a random variable. If the estimator is derived from a sample statistic and is used to estimate some population statistic, then the expectation is with respect to the sampling distribution of the sample statistic.

The MSE can be written as the sum of the [variance](https://en.wikipedia.org/wiki/Variance) of the estimator and the squared [bias](https://en.wikipedia.org/wiki/Bias_%28statistics%29) of the estimator, providing a useful way to calculate the MSE and implying that in the case of unbiased estimators, the MSE and variance are equivalent.[[2]](https://en.wikipedia.org/wiki/Mean_squared_error#cite_note-wackerly-2)

MSE ⁡ ( θ ^ ) = Var ⁡ ( θ ^ ) + ( Bias ⁡ ( θ ^ , θ ) ) 2 . {\displaystyle \operatorname {MSE} ({\hat {\theta }})=\operatorname {Var} ({\hat {\theta }})+\left(\operatorname {Bias} ({\hat {\theta }},\theta )\right)^{2}.}

**Proof of variance and bias relationship**

MSE ⁡ ( θ ^ ) ≡ E ( ( θ ^ − θ ) 2 ) = E [ ( θ ^ − E ( θ ^ ) + E ( θ ^ ) − θ ) 2 ] = E [ ( θ ^ − E ( θ ^ ) ) 2 + 2 ( ( θ ^ − E ( θ ^ ) ) ( E ( θ ^ ) − θ ) ) + ( E ( θ ^ ) − θ ) 2 ] = E [ ( θ ^ − E ( θ ^ ) ) 2 ] + 2 E [ ( θ ^ − E ( θ ^ ) ) ( E ( θ ^ ) − θ ⏞ This is a constant, so it can be pulled out. ) ] + E [ ( E ( θ ^ ) − θ ) 2 ⏞ This is a constant, so its expected value is itself. ] = E [ ( θ ^ − E ( θ ^ ) ) 2 ] + 2 E ( θ ^ − E ( θ ^ ) ) ⏟ = E ( θ ^ ) − E ( θ ^ ) = 0 ( E ( θ ^ ) − θ ) + ( E ( θ ^ ) − θ ) 2 = E [ ( θ ^ − E ( θ ^ ) ) 2 ] + ( E ( θ ^ ) − θ ) 2 = Var ⁡ ( θ ^ ) + Bias ⁡ ( θ ^ , θ ) 2 {\displaystyle {\begin{aligned}\operatorname {MSE} ({\hat {\theta }})\equiv \mathbb {E} (({\hat {\theta }}-\theta )^{2})&=\mathbb {E} \left[\left({\hat {\theta }}-\mathbb {E} ({\hat {\theta }})+\mathbb {E} ({\hat {\theta }})-\theta \right)^{2}\right]\\&=\mathbb {E} \left[\left({\hat {\theta }}-\mathbb {E} ({\hat {\theta }})\right)^{2}+2\left(({\hat {\theta }}-\mathbb {E} ({\hat {\theta }}))(\mathbb {E} ({\hat {\theta }})-\theta )\right)+\left(\mathbb {E} ({\hat {\theta }})-\theta \right)^{2}\right]\\&=\mathbb {E} \left[\left({\hat {\theta }}-\mathbb {E} ({\hat {\theta }})\right)^{2}\right]+2\mathbb {E} {\Big [}({\hat {\theta }}-\mathbb {E} ({\hat {\theta }}))(\overbrace {\mathbb {E} ({\hat {\theta }})-\theta } ^{\begin{smallmatrix}{\text{This is}}\\{\text{a constant,}}\\{\text{so it can be}}\\{\text{pulled out.}}\end{smallmatrix}})\,{\Big ]}+\mathbb {E} {\Big [}\,\overbrace {\left(\mathbb {E} ({\hat {\theta }})-\theta \right)^{2}} ^{\begin{smallmatrix}{\text{This is a}}\\{\text{constant, so its}}\\{\text{expected value}}\\{\text{is itself.}}\end{smallmatrix}}\,{\Big ]}\\&=\mathbb {E} \left[\left({\hat {\theta }}-\mathbb {E} ({\hat {\theta }})\right)^{2}\right]+2\underbrace {\mathbb {E} ({\hat {\theta }}-\mathbb {E} ({\hat {\theta }}))} \_{=\mathbb {E} ({\hat {\theta }})-\mathbb {E} ({\hat {\theta }})=0}(\mathbb {E} ({\hat {\theta }})-\theta )+\left(\mathbb {E} ({\hat {\theta }})-\theta \right)^{2}\\&=\mathbb {E} \left[\left({\hat {\theta }}-\mathbb {E} ({\hat {\theta }})\right)^{2}\right]+\left(\mathbb {E} ({\hat {\theta }})-\theta \right)^{2}\\&=\operatorname {Var} ({\hat {\theta }})+\operatorname {Bias} ({\hat {\theta }},\theta )^{2}\end{aligned}}}

**Regression**

In regression analysis, the term *mean squared error* is sometimes used to refer to the unbiased estimate of error variance: the [residual sum of squares](https://en.wikipedia.org/wiki/Residual_sum_of_squares) divided by the number of [degrees of freedom](https://en.wikipedia.org/wiki/Degrees_of_freedom_%28statistics%29). This definition for a known, computed quantity differs from the above definition for the computed MSE of a predictor in that a different denominator is used. The denominator is the sample size reduced by the number of model parameters estimated from the same data, *(n-p)* for *p* [regressors](https://en.wikipedia.org/wiki/Regressor) or *(n-p-1)* if an intercept is used.[[3]](https://en.wikipedia.org/wiki/Mean_squared_error#cite_note-3) For more details, see [errors and residuals in statistics](https://en.wikipedia.org/wiki/Errors_and_residuals_in_statistics). Note that, although the MSE is not an unbiased estimator of the error variance, it is consistent, given the consistency of the predictor.

Also in regression analysis, "mean squared error", often referred to as [mean squared prediction error](https://en.wikipedia.org/wiki/Mean_squared_prediction_error) or "out-of-sample mean squared error", can refer to the mean value of the squared deviations of the predictions from the true values, over an out-of-sample test space, generated by a model estimated over a particular sample space. This also is a known, computed quantity, and it varies by sample and by out-of-sample test space.

**Examples**

**Mean**

Suppose we have a random sample of size *n* from a population, X 1 , … , X n {\displaystyle X\_{1},\dots ,X\_{n}} . Suppose the sample units were chosen with replacement. That is, the *n* units are selected one at a time, and previously selected units are still eligible for selection for all *n* draws. The usual estimator for the mean is the sample average

X ¯ = 1 n ∑ i = 1 n X i {\displaystyle {\overline {X}}={\frac {1}{n}}\sum \_{i=1}^{n}X\_{i}}

which has an expected value equal to the true mean μ (so it is unbiased) and a mean square error of

MSE ⁡ ( X ¯ ) = E ⁡ [ ( X ¯ − μ ) 2 ] = ( σ n ) 2 = σ 2 n {\displaystyle \operatorname {MSE} \left({\overline {X}}\right)=\operatorname {E} \left[\left({\overline {X}}-\mu \right)^{2}\right]=\left({\frac {\sigma }{\sqrt {n}}}\right)^{2}={\frac {\sigma ^{2}}{n}}}

where σ 2 {\displaystyle \sigma ^{2}} is the [population variance](https://en.wikipedia.org/wiki/Sample_variance#Population_variance).

For a [Gaussian distribution](https://en.wikipedia.org/wiki/Gaussian_distribution) this is the [best unbiased estimator](https://en.wikipedia.org/wiki/Best_unbiased_estimator) (that is, it has the lowest MSE among all unbiased estimators), but not, say, for a [uniform distribution](https://en.wikipedia.org/wiki/Uniform_distribution_%28continuous%29).

**Variance**

Further information: [Sample variance](https://en.wikipedia.org/wiki/Sample_variance)

The usual estimator for the variance is the *corrected* [*sample variance*](https://en.wikipedia.org/wiki/Sample_variance)*:*

S n − 1 2 = 1 n − 1 ∑ i = 1 n ( X i − X ¯ ) 2 = 1 n − 1 ( ∑ i = 1 n X i 2 − n X ¯ 2 ) . {\displaystyle S\_{n-1}^{2}={\frac {1}{n-1}}\sum \_{i=1}^{n}\left(X\_{i}-{\overline {X}}\,\right)^{2}={\frac {1}{n-1}}\left(\sum \_{i=1}^{n}X\_{i}^{2}-n{\overline {X}}^{2}\right).}

This is unbiased (its expected value is σ 2 {\displaystyle \sigma ^{2}} ), hence also called the *unbiased sample variance,* and its MSE is[[4]](https://en.wikipedia.org/wiki/Mean_squared_error#cite_note-4)

MSE ⁡ ( S n − 1 2 ) = 1 n ( μ 4 − n − 3 n − 1 σ 4 ) = 1 n ( γ 2 + 2 n n − 1 ) σ 4 , {\displaystyle {\begin{aligned}\operatorname {MSE} (S\_{n-1}^{2})&={\frac {1}{n}}\left(\mu \_{4}-{\frac {n-3}{n-1}}\sigma ^{4}\right)\\&={\frac {1}{n}}\left(\gamma \_{2}+{\frac {2n}{n-1}}\right)\sigma ^{4},\end{aligned}}}

where μ 4 {\displaystyle \mu \_{4}} is the fourth [central moment](https://en.wikipedia.org/wiki/Central_moment) of the distribution or population and γ 2 = μ 4 / σ 4 − 3 {\displaystyle \gamma \_{2}=\mu \_{4}/\sigma ^{4}-3} is the [excess kurtosis](https://en.wikipedia.org/wiki/Excess_kurtosis).

However, one can use other estimators for σ 2 {\displaystyle \sigma ^{2}} which are proportional to S n − 1 2 {\displaystyle S\_{n-1}^{2}} , and an appropriate choice can always give a lower mean square error. If we define

S a 2 = n − 1 a S n − 1 2 = 1 a ∑ i = 1 n ( X i − X ¯ ) 2 {\displaystyle {\begin{aligned}S\_{a}^{2}&={\frac {n-1}{a}}S\_{n-1}^{2}\\&={\frac {1}{a}}\sum \_{i=1}^{n}\left(X\_{i}-{\overline {X}}\,\right)^{2}\end{aligned}}}

then the MSE is

MSE ⁡ ( S a 2 ) = E ⁡ ( ( n − 1 a S n − 1 2 − σ 2 ) 2 ) = n − 1 n a 2 [ ( n − 1 ) γ 2 + n 2 + n ] σ 4 − 2 ( n − 1 ) a σ 4 + σ 4 {\displaystyle {\begin{aligned}\operatorname {MSE} (S\_{a}^{2})&=\operatorname {E} \left(\left({\frac {n-1}{a}}S\_{n-1}^{2}-\sigma ^{2}\right)^{2}\right)\\&={\frac {n-1}{na^{2}}}[(n-1)\gamma \_{2}+n^{2}+n]\sigma ^{4}-{\frac {2(n-1)}{a}}\sigma ^{4}+\sigma ^{4}\end{aligned}}}

This is minimized when

a = ( n − 1 ) γ 2 + n 2 + n n = n + 1 + n − 1 n γ 2 . {\displaystyle a={\frac {(n-1)\gamma \_{2}+n^{2}+n}{n}}=n+1+{\frac {n-1}{n}}\gamma \_{2}.}

For a [Gaussian distribution](https://en.wikipedia.org/wiki/Gaussian_distribution), where γ 2 = 0 {\displaystyle \gamma \_{2}=0} , this means the MSE is minimized when dividing the sum by a = n + 1 {\displaystyle a=n+1} . The minimum excess kurtosis is γ 2 = − 2 {\displaystyle \gamma \_{2}=-2} ,[[a]](https://en.wikipedia.org/wiki/Mean_squared_error#cite_note-5) which is achieved by a [Bernoulli distribution](https://en.wikipedia.org/wiki/Bernoulli_distribution) with *p* = 1/2 (a coin flip), and the MSE is minimized for a = n − 1 + 2 / n {\displaystyle a=n-1+2/n} . So no matter what the kurtosis, we get a "better" estimate (in the sense of having a lower MSE) by scaling down the unbiased estimator a little bit; this is a simple example of a [shrinkage estimator](https://en.wikipedia.org/wiki/Shrinkage_estimator): one "shrinks" the estimator towards zero (scales down the unbiased estimator).

Further, while the corrected sample variance is the [best unbiased estimator](https://en.wikipedia.org/wiki/Best_unbiased_estimator) (minimum mean square error among unbiased estimators) of variance for Gaussian distributions, if the distribution is not Gaussian then even among unbiased estimators, the best unbiased estimator of the variance may not be S n − 1 2 . {\displaystyle S\_{n-1}^{2}.}

**Gaussian distribution**

The following table gives several estimators of the true parameters of the population, μ and σ2, for the Gaussian case.[[5]](https://en.wikipedia.org/wiki/Mean_squared_error#cite_note-6)

|  |  |  |
| --- | --- | --- |
| **True value** | **Estimator** | **Mean squared error** |
| θ = μ | θ ^ {\displaystyle {\hat {\theta }}} = the unbiased estimator of the [population mean](https://en.wikipedia.org/wiki/Population_mean), X ¯ = 1 n ∑ i = 1 n ( X i ) {\displaystyle {\overline {X}}={\frac {1}{n}}\sum \_{i=1}^{n}(X\_{i})} | MSE ⁡ ( X ¯ ) = E ⁡ ( ( X ¯ − μ ) 2 ) = ( σ n ) 2 {\displaystyle \operatorname {MSE} ({\overline {X}})=\operatorname {E} (({\overline {X}}-\mu )^{2})=\left({\frac {\sigma }{\sqrt {n}}}\right)^{2}} |
| θ = σ2 | θ ^ {\displaystyle {\hat {\theta }}} = the unbiased estimator of the [population variance](https://en.wikipedia.org/wiki/Population_variance), S n − 1 2 = 1 n − 1 ∑ i = 1 n ( X i − X ¯ ) 2 {\displaystyle S\_{n-1}^{2}={\frac {1}{n-1}}\sum \_{i=1}^{n}\left(X\_{i}-{\overline {X}}\,\right)^{2}} | MSE ⁡ ( S n − 1 2 ) = E ⁡ ( ( S n − 1 2 − σ 2 ) 2 ) = 2 n − 1 σ 4 {\displaystyle \operatorname {MSE} (S\_{n-1}^{2})=\operatorname {E} ((S\_{n-1}^{2}-\sigma ^{2})^{2})={\frac {2}{n-1}}\sigma ^{4}} |
| θ = σ2 | θ ^ {\displaystyle {\hat {\theta }}} = the biased estimator of the [population variance](https://en.wikipedia.org/wiki/Population_variance), S n 2 = 1 n ∑ i = 1 n ( X i − X ¯ ) 2 {\displaystyle S\_{n}^{2}={\frac {1}{n}}\sum \_{i=1}^{n}\left(X\_{i}-{\overline {X}}\,\right)^{2}} | MSE ⁡ ( S n 2 ) = E ⁡ ( ( S n 2 − σ 2 ) 2 ) = 2 n − 1 n 2 σ 4 {\displaystyle \operatorname {MSE} (S\_{n}^{2})=\operatorname {E} ((S\_{n}^{2}-\sigma ^{2})^{2})={\frac {2n-1}{n^{2}}}\sigma ^{4}} |
| θ = σ2 | θ ^ {\displaystyle {\hat {\theta }}} = the biased estimator of the [population variance](https://en.wikipedia.org/wiki/Population_variance), S n + 1 2 = 1 n + 1 ∑ i = 1 n ( X i − X ¯ ) 2 {\displaystyle S\_{n+1}^{2}={\frac {1}{n+1}}\sum \_{i=1}^{n}\left(X\_{i}-{\overline {X}}\,\right)^{2}} | MSE ⁡ ( S n + 1 2 ) = E ⁡ ( ( S n + 1 2 − σ 2 ) 2 ) = 2 n + 1 σ 4 {\displaystyle \operatorname {MSE} (S\_{n+1}^{2})=\operatorname {E} ((S\_{n+1}^{2}-\sigma ^{2})^{2})={\frac {2}{n+1}}\sigma ^{4}} |

Note that:

1. The MSEs shown for the variance estimators assume X i ∼ N ⁡ ( μ , σ 2 ) {\displaystyle X\_{i}\sim \operatorname {N} (\mu ,\sigma ^{2})} [i.i.d.](https://en.wikipedia.org/wiki/Independent_and_identically-distributed_random_variables) so that ( n − 1 ) S n − 1 2 σ 2 ∼ χ n − 1 2 {\displaystyle {\frac {(n-1)S\_{n-1}^{2}}{\sigma ^{2}}}\sim \chi \_{n-1}^{2}} . The result for S n − 1 2 {\displaystyle S\_{n-1}^{2}} follows easily from the χ n − 1 2 {\displaystyle \chi \_{n-1}^{2}} variance that is 2 n − 2 {\displaystyle 2n-2} .
2. Unbiased estimators may not produce estimates with the smallest total variation (as measured by MSE): the MSE of S n − 1 2 {\displaystyle S\_{n-1}^{2}} is larger than that of S n + 1 2 {\displaystyle S\_{n+1}^{2}} or S n 2 {\displaystyle S\_{n}^{2}} .
3. Estimators with the smallest total variation may produce biased estimates: S n + 1 2 {\displaystyle S\_{n+1}^{2}} typically underestimates σ2 by 2 n σ 2 {\displaystyle {\frac {2}{n}}\sigma ^{2}}

**Interpretation**

An MSE of zero, meaning that the estimator θ ^ {\displaystyle {\hat {\theta }}} predicts observations of the parameter θ {\displaystyle \theta } with perfect accuracy, is the ideal, but is typically not possible.

Values of MSE may be used for comparative purposes. Two or more [statistical models](https://en.wikipedia.org/wiki/Statistical_model) may be compared using their MSEs as a measure of how well they explain a given set of observations: An unbiased estimator (estimated from a statistical model) with the smallest variance among all unbiased estimators is the [best unbiased estimator]] or MVUE (Minimum Variance Unbiased Estimator).

Both [linear regression](https://en.wikipedia.org/wiki/Linear_regression) techniques such as [analysis of variance](https://en.wikipedia.org/wiki/Analysis_of_variance) estimate the MSE as part of the analysis and use the estimated MSE to determine the [statistical significance](https://en.wikipedia.org/wiki/Statistical_significance) of the factors or predictors under study. The goal of [experimental design](https://en.wikipedia.org/wiki/Experimental_design) is to construct experiments in such a way that when the observations are analyzed, the MSE is close to zero relative to the magnitude of at least one of the estimated treatment effects.

MSE is also used in several [stepwise regression](https://en.wikipedia.org/wiki/Stepwise_regression) techniques as part of the determination as to how many predictors from a candidate set to include in a model for a given set of observations.

**Applications**

* Minimizing MSE is a key criterion in selecting estimators: see [minimum mean-square error](https://en.wikipedia.org/wiki/Minimum_mean-square_error). Among unbiased estimators, minimizing the MSE is equivalent to minimizing the variance, and the estimator that does this is the [minimum variance unbiased estimator](https://en.wikipedia.org/wiki/Minimum_variance_unbiased_estimator). However, a biased estimator may have lower MSE; see [estimator bias](https://en.wikipedia.org/wiki/Estimator_bias).
* In [statistical modelling](https://en.wikipedia.org/wiki/Statistical_modelling) the MSE, representing the difference between the actual observations and the observation values predicted by the model, is used to determine the extent to which the model fits the data and whether the removal or some explanatory variables, simplifying the model, is possible without significantly harming the model's predictive ability.

**Loss function**

Squared error loss is one of the most widely used [loss functions](https://en.wikipedia.org/wiki/Loss_function) in statistics, though its widespread use stems more from mathematical convenience than considerations of actual loss in applications. [Carl Friedrich Gauss](https://en.wikipedia.org/wiki/Carl_Friedrich_Gauss), who introduced the use of mean squared error, was aware of its arbitrariness and was in agreement with objections to it on these grounds.[[1]](https://en.wikipedia.org/wiki/Mean_squared_error#cite_note-pointEstimation-1) The mathematical benefits of mean squared error are particularly evident in its use at analyzing the performance of [linear regression](https://en.wikipedia.org/wiki/Linear_regression), as it allows one to partition the variation in a dataset into variation explained by the model and variation explained by randomness.

**Criticism**

The use of mean squared error without question has been criticized by the [decision theorist](https://en.wikipedia.org/wiki/Decision_theory) [James Berger](https://en.wikipedia.org/wiki/James_Berger_%28statistician%29). Mean squared error is the negative of the expected value of one specific [utility function](https://en.wikipedia.org/wiki/Utility_function), the quadratic utility function, which may not be the appropriate utility function to use under a given set of circumstances. There are, however, some scenarios where mean squared error can serve as a good approximation to a loss function occurring naturally in an application.[[6]](https://en.wikipedia.org/wiki/Mean_squared_error#cite_note-7)

Like [variance](https://en.wikipedia.org/wiki/Variance), mean squared error has the disadvantage of heavily weighting [outliers](https://en.wikipedia.org/wiki/Outliers).[[7]](https://en.wikipedia.org/wiki/Mean_squared_error#cite_note-8) This is a result of the squaring of each term, which effectively weights large errors more heavily than small ones. This property, undesirable in many applications, has led researchers to use alternatives such as the [mean absolute error](https://en.wikipedia.org/wiki/Mean_absolute_error), or those based on the [median](https://en.wikipedia.org/wiki/Median).

**See also**

* [James–Stein estimator](https://en.wikipedia.org/wiki/James%E2%80%93Stein_estimator)
* [Hodges' estimator](https://en.wikipedia.org/wiki/Hodges%27_estimator)
* [Mean percentage error](https://en.wikipedia.org/wiki/Mean_percentage_error)
* [Mean square weighted deviation](https://en.wikipedia.org/wiki/Mean_square_weighted_deviation)
* [Mean squared displacement](https://en.wikipedia.org/wiki/Mean_squared_displacement)
* [Mean squared prediction error](https://en.wikipedia.org/wiki/Mean_squared_prediction_error)
* [Minimum mean squared error estimator](https://en.wikipedia.org/wiki/Minimum_mean_squared_error)
* [Peak signal-to-noise ratio](https://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio)
* [Root mean square deviation](https://en.wikipedia.org/wiki/Root_mean_square_deviation)
* [Squared deviations](https://en.wikipedia.org/wiki/Squared_deviations)

**Notes**

* 1. This can be proved by [Jensen's inequality](https://en.wikipedia.org/wiki/Jensen%27s_inequality) as follows. The fourth [central moment](https://en.wikipedia.org/wiki/Central_moment) is an upper bound for the square of variance, so that the least value for their ratio is one, therefore, the least value for the [excess kurtosis](https://en.wikipedia.org/wiki/Excess_kurtosis) is −2, achieved, for instance, by a Bernoulli with p=1/2.

**References**

 *Lehmann, E. L.; Casella, George (1998). Theory of Point Estimation (2nd ed.). New York: Springer.* [*ISBN*](https://en.wikipedia.org/wiki/International_Standard_Book_Number)[*0-387-98502-6*](https://en.wikipedia.org/wiki/Special:BookSources/0-387-98502-6)*.* [*MR*](https://en.wikipedia.org/wiki/Mathematical_Reviews)[*1639875*](https://www.ams.org/mathscinet-getitem?mr=1639875)*.*

  *Wackerly, Dennis; Mendenhall, William; Scheaffer, Richard L. (2008). Mathematical Statistics with Applications (7 ed.). Belmont, CA, USA: Thomson Higher Education.* [*ISBN*](https://en.wikipedia.org/wiki/International_Standard_Book_Number)[*0-495-38508-5*](https://en.wikipedia.org/wiki/Special:BookSources/0-495-38508-5)*.*

  Steel, R.G.D, and Torrie, J. H., *Principles and Procedures of Statistics with Special Reference to the Biological Sciences.*, [McGraw Hill](https://en.wikipedia.org/wiki/McGraw_Hill), 1960, page 288.

  *Mood, A.; Graybill, F.; Boes, D. (1974). Introduction to the Theory of Statistics (3rd ed.). McGraw-Hill. p. 229.*

  [*DeGroot, Morris H.*](https://en.wikipedia.org/wiki/Morris_H._DeGroot) *(1980). Probability and Statistics (2nd ed.). Addison-Wesley.*

  [*Berger, James O.*](https://en.wikipedia.org/wiki/James_Berger_%28statistician%29) *(1985). "2.4.2 Certain Standard Loss Functions". Statistical decision theory and Bayesian Analysis (2nd ed.). New York: Springer-Verlag. p. 60.* [*ISBN*](https://en.wikipedia.org/wiki/International_Standard_Book_Number)[*0-387-96098-8*](https://en.wikipedia.org/wiki/Special:BookSources/0-387-96098-8)*.* [*MR*](https://en.wikipedia.org/wiki/Mathematical_Reviews)[*0804611*](https://www.ams.org/mathscinet-getitem?mr=0804611)*.*

* 1.  Sergio Bermejo, Joan Cabestany (2001) "[Oriented principal component analysis for large margin classifiers](http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6T08-43PS3GC-1&_user=483692&_coverDate=12%2F31%2F2001&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000022720&_version=1&_urlVersion=0&_userid=483692&md5=8586e409694e1b50da3aa3c6fce18cb8)", *Neural Networks*, 14 (10), 1447–1461.

[Categories](https://en.wikipedia.org/wiki/Help:Category):

* [Estimation theory](https://en.wikipedia.org/wiki/Category:Estimation_theory)
* [Point estimation performance](https://en.wikipedia.org/wiki/Category:Point_estimation_performance)
* [Statistical deviation and dispersion](https://en.wikipedia.org/wiki/Category:Statistical_deviation_and_dispersion)
* [Loss functions](https://en.wikipedia.org/wiki/Category:Loss_functions)
* [Least squares](https://en.wikipedia.org/wiki/Category:Least_squares)

**Navigation menu**

* Not logged in
* [Talk](https://en.wikipedia.org/wiki/Special:MyTalk)
* [Contributions](https://en.wikipedia.org/wiki/Special:MyContributions)
* [Create account](https://en.wikipedia.org/w/index.php?title=Special:CreateAccount&returnto=Mean+squared+error)
* [Log in](https://en.wikipedia.org/w/index.php?title=Special:UserLogin&returnto=Mean+squared+error)
* [Article](https://en.wikipedia.org/wiki/Mean_squared_error)
* [Talk](https://en.wikipedia.org/wiki/Talk:Mean_squared_error)
* [Read](https://en.wikipedia.org/wiki/Mean_squared_error)
* [Edit](https://en.wikipedia.org/w/index.php?title=Mean_squared_error&action=edit)
* [View history](https://en.wikipedia.org/w/index.php?title=Mean_squared_error&action=history)

Top of Form

Bottom of Form

Top of Form

Bottom of Form

* [Main page](https://en.wikipedia.org/wiki/Main_Page)
* [Contents](https://en.wikipedia.org/wiki/Portal:Contents)
* [Featured content](https://en.wikipedia.org/wiki/Portal:Featured_content)
* [Current events](https://en.wikipedia.org/wiki/Portal:Current_events)
* [Random article](https://en.wikipedia.org/wiki/Special:Random)
* [Donate to Wikipedia](https://donate.wikimedia.org/wiki/Special:FundraiserRedirector?utm_source=donate&utm_medium=sidebar&utm_campaign=C13_en.wikipedia.org&uselang=en)
* [Wikipedia store](https://shop.wikimedia.org)

**Interaction**

* [Help](https://en.wikipedia.org/wiki/Help:Contents)
* [About Wikipedia](https://en.wikipedia.org/wiki/Wikipedia:About)
* [Community portal](https://en.wikipedia.org/wiki/Wikipedia:Community_portal)
* [Recent changes](https://en.wikipedia.org/wiki/Special:RecentChanges)
* [Contact page](https://en.wikipedia.org/wiki/Wikipedia:Contact_us)

**Tools**

* [What links here](https://en.wikipedia.org/wiki/Special:WhatLinksHere/Mean_squared_error)
* [Related changes](https://en.wikipedia.org/wiki/Special:RecentChangesLinked/Mean_squared_error)
* [Upload file](https://en.wikipedia.org/wiki/Wikipedia:File_Upload_Wizard)
* [Special pages](https://en.wikipedia.org/wiki/Special:SpecialPages)
* [Permanent link](https://en.wikipedia.org/w/index.php?title=Mean_squared_error&oldid=731486610)
* [Page information](https://en.wikipedia.org/w/index.php?title=Mean_squared_error&action=info)
* [Wikidata item](https://www.wikidata.org/wiki/Q1940696)
* [Cite this page](https://en.wikipedia.org/w/index.php?title=Special:CiteThisPage&page=Mean_squared_error&id=731486610)

**Print/export**

* [Create a book](https://en.wikipedia.org/w/index.php?title=Special:Book&bookcmd=book_creator&referer=Mean+squared+error)
* [Download as PDF](https://en.wikipedia.org/w/index.php?title=Special:Book&bookcmd=render_article&arttitle=Mean+squared+error&returnto=Mean+squared+error&oldid=731486610&writer=rdf2latex)
* [Printable version](https://en.wikipedia.org/w/index.php?title=Mean_squared_error&printable=yes)

**Languages**

* [العربية](https://ar.wikipedia.org/wiki/%D8%AE%D8%B7%D8%A3_%D8%AA%D8%B1%D8%A8%D9%8A%D8%B9%D9%8A_%D9%85%D8%AA%D9%88%D8%B3%D8%B7)
* [Català](https://ca.wikipedia.org/wiki/Error_quadr%C3%A0tic_mig)
* [Deutsch](https://de.wikipedia.org/wiki/Mittlere_quadratische_Abweichung)
* [Español](https://es.wikipedia.org/wiki/Error_cuadr%C3%A1tico_medio)
* [Euskara](https://eu.wikipedia.org/wiki/Batez_besteko_errore_koadratiko)
* [Français](https://fr.wikipedia.org/wiki/Erreur_quadratique_moyenne)
* [Italiano](https://it.wikipedia.org/wiki/Errore_quadratico_medio)
* [עברית](https://he.wikipedia.org/wiki/%D7%98%D7%A2%D7%95%D7%AA_%D7%A8%D7%99%D7%91%D7%95%D7%A2%D7%99%D7%AA_%D7%9E%D7%9E%D7%95%D7%A6%D7%A2%D7%AA)
* [Polski](https://pl.wikipedia.org/wiki/B%C5%82%C4%85d_%C5%9Bredniokwadratowy)
* [Português](https://pt.wikipedia.org/wiki/Erro_quadr%C3%A1tico_m%C3%A9dio)
* [Basa Sunda](https://su.wikipedia.org/wiki/Mean_kuadrat_kasalahan)
* [Tiếng Việt](https://vi.wikipedia.org/wiki/Sai_s%E1%BB%91_to%C3%A0n_ph%C6%B0%C6%A1ng_trung_b%C3%ACnh)
* [中文](https://zh.wikipedia.org/wiki/%E5%9D%87%E6%96%B9%E5%B7%AE)

[Edit links](https://www.wikidata.org/wiki/Q1940696#sitelinks-wikipedia)

* This page was last modified on 25 July 2016, at 17:46.
* Text is available under the [Creative Commons Attribution-ShareAlike License](https://en.wikipedia.org/wiki/Wikipedia:Text_of_Creative_Commons_Attribution-ShareAlike_3.0_Unported_License); additional terms may apply. By using this site, you agree to the [Terms of Use](https://wikimediafoundation.org/wiki/Terms_of_Use) and [Privacy Policy](https://wikimediafoundation.org/wiki/Privacy_policy). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](https://www.wikimediafoundation.org/), a non-profit organization.
* [Privacy policy](https://wikimediafoundation.org/wiki/Privacy_policy)
* [About Wikipedia](https://en.wikipedia.org/wiki/Wikipedia:About)
* [Disclaimers](https://en.wikipedia.org/wiki/Wikipedia:General_disclaimer)
* [Contact Wikipedia](https://en.wikipedia.org/wiki/Wikipedia:Contact_us)
* [Developers](https://www.mediawiki.org/wiki/Special:MyLanguage/How_to_contribute)
* [Cookie statement](https://wikimediafoundation.org/wiki/Cookie_statement)
* [Mobile view](https://en.m.wikipedia.org/w/index.php?title=Mean_squared_error&mobileaction=toggle_view_mobile)
* [Wikimedia Foundation](https://wikimediafoundation.org/)
* [Powered by MediaWiki](https://www.mediawiki.org/)