# What is the coefficient of variation?

Authored by
**stats writer**

June 30, 2024

The coefficient of variation (CV) is a statistical measure that describes the ratio of the standard deviation to the mean of a dataset. It is used to assess the relative variability of a dataset and is often expressed as a percentage. A lower CV indicates a lower variability and a more homogeneous dataset, while a higher CV indicates a higher variability and a more heterogeneous dataset. The CV is commonly used in fields such as finance, economics, and science to compare the variability of different datasets and make informed decisions based on the results.

## FAQ: What is the coefficient of variation?

## FAQ: What is the coefficient of variation?

### Situations and Definitions

**A coefficient of variation (CV) can be calculated and interpreted in two different settings: analyzing a single variable and interpreting a model. The standard formulation of the CV, the ratio of the standard deviation to the mean, applies in the single variable setting. In the modeling setting, the CV is calculated as the ratio of the root mean squared error (RMSE) to the mean of the dependent variable. In both settings, the CV is often presented as the given ratio multiplied by 100. The CV for a single variable aims to describe the dispersion of the variable**

**in a way that does not depend on the variable's measurement unit.**
**The higher the CV, the greater the dispersion in the variable. The CV for a model aims to describe the model fit**
**in terms of the relative sizes of the squared residuals and outcome values. The**
**lower the CV, the smaller the residuals relative to the predicted value.**
**This is suggestive of a good model fit.**

**The CV for a variable can easily be calculated using the information from a**
**typical variable summary (and sometimes the CV will be returned by default in**
**the variable summary). We demonstrate below how to calculate the CV in**
**Stata.**

**use https://stats.idre.ucla.edu/stat/stata/notes/hsb1, clear**
**summarize math**

**Variable | Obs Mean Std. Dev. Min Max**
**-------------+---------------------------------------------------------**

**math | 200 52.645 9.368448 33 75**

**di 100 * r(sd) / r(mean)**

**17.795513**

**The CV for a model can similarly be calculated when it is not included in the model output.**

**regress math socst**

**Source | SS df MS Number of obs = 200**
**-------------+------------------------------ F( 1, 198) = 83.43**
**Model | 5177.88866 1 5177.88866 Prob > F = 0.0000**
**Residual | 12287.9063 198 62.060133 R-squared = 0.2965**
**-------------+------------------------------ Adj R-squared = 0.2929**
**Total | 17465.795 199 87.7678141 Root MSE = 7.8778**

**------------------------------------------------------------------------------**
**math | Coef. Std. Err. t P>|t|**
**-------------+----------------------------------------------------------------**
**socst | .4751335 .052017 9.13 0.000 .372555 .577712**
**_cons | 27.74563 2.782287 9.97 0.000 22.25891 33.23235**
**------------------------------------------------------------------------------**

**quietly summarize math**
**di 100 * e(rmse) / r(mean)**

**14.964052**

### Advantages

**The advantage of the CV is that it is unitless. This allows CVs to be compared to each other in ways that other measures, like standard deviations or root mean squared residuals, cannot be.**

**In the variable CV setting: The standard deviations of two variables, while both measure dispersion in their respective variables, cannot be compared to each other in a meaningful way to determine which variable has greater dispersion because they may vary greatly in their units and the means about which they occur. The standard deviation and mean of a variable are expressed in the same units, so taking the ratio of these two**

**allows the units to cancel. This ratio can then be compared to other such ratios in a meaningful way: between two variables (that meet the assumptions outlined below), the variable with the smaller CV is less dispersed than the variable with the larger CV.**

**In the model CV setting: Similarly, the RMSE of two models both measure the magnitude of the residuals, but they cannot be compared to each other in a meaningful way to determine which model provides better predictions of an outcome. The model RMSE and mean of the predicted variable are expressed in the same units, so taking the ratio of these two allows the units to cancel. This ratio can then be compared to other such ratios in a meaningful way: between two models (where the outcome variable meets the assumptions outlined below), the model with the smaller CV has predicted values that are closer to the actual values. It is**

**interesting to note the differences between a model's CV and R-squared values. Both are unitless measures that are indicative of model fit, but they define model fit in two different ways: CV evaluates the relative closeness of the predictions to the actual values while R-squared evaluates how much of the variability in the actual values is explained by the model.**

### Requirements and Disadvantages

**There are some requirements that must be met in order for the CV to be interpreted in the ways we have described. The most obvious problem arises when the mean of a variable is zero. In this case, the CV cannot be calculated. Even if the mean of a variable is not zero, but the variable contains both positive and negative values and the mean is close to zero, then the CV can be misleading. The CV of a variable or the CV of a prediction**

**model for a variable can be considered as a reasonable measure if the variable**
**contains only positive values. This is a definite disadvantage of CVs.**