

Source: <https://www.investopedia.com/terms/m/multicollinearity.asp>

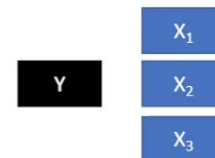
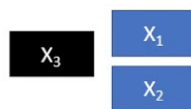
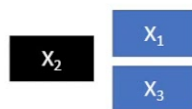
Source: <https://www.investopedia.com/terms/v/variance-inflation-factor.asp>

## MULTICOLLINEARITY AND VARIANCE INFLATION FACTOR (VIF):

### What Is Multicollinearity?

#### Multicollinearity

- Increase difficulty in estimation of coefficients
- Inflates coefficients
- Variance inflation factor (VIF) – a number to quantify the inflation seen in the coefficients due to multicollinearity
- $VIF = \frac{1}{1 - R_{sq}^2}$
- VIF = 1 means no collinearity (orthogonal)
- VIF between 5-10 or higher indicates collinearity
- **Solution:** Find VIF, remove the redundant term



R <sup>2</sup>	VIF
<0.9	1
0.9	10
0.99	100

Multicollinearity is the occurrence of high intercorrelations among two or more independent variables in a multiple regression model. Multicollinearity can lead to skewed or misleading results when a researcher or analyst attempts to determine how well each independent variable can be used most effectively to predict or understand the dependent variable in a statistical model.

In general, multicollinearity can lead to wider confidence intervals that produce less reliable probabilities in terms of the effect of independent variables in a model.

In technical analysis, multicollinearity can lead to incorrect assumptions about an investment. It generally occurs because multiple indicators of the same type have been used to analyze a stock.

## KEY TAKEAWAYS

- Multicollinearity is a statistical concept where several independent variables in a model are correlated.
- Two variables are considered perfectly collinear if their correlation coefficient is  $\pm 1.0$ .
- Multicollinearity among independent variables will result in less reliable statistical inferences.
- When you're analyzing an investment, it is better to use different types of indicators rather than multiple indicators of the same type to avoid multicollinearity.
- Multicollinearity can lead to less reliable results because the results you're comparing are generally the same.

## Understanding Multicollinearity

Statistical analysts use [multiple regression models](#) to predict the value of a specified dependent variable based on the values of two or more independent variables. The dependent variable is sometimes called the outcome, target, or criterion variable.

An example is a [multivariate regression model](#) that attempts to anticipate stock returns based on metrics such as the [price-to-earnings ratio](#) (P/E ratios), market capitalization, or other data. The stock return is the dependent variable (the outcome), and the various bits of financial data are the independent variables.

Multicollinearity in a multiple regression model indicates that collinear independent variables are not truly independent. For example, past performance might be related to [market capitalization](#). The stocks of businesses that have performed well experience investor confidence, increasing demand for that company's stock, which increases its market value.

### Effects of Multicollinearity

Although multicollinearity does not affect the regression estimates, it makes them vague, imprecise, and unreliable. Thus, it can be hard to determine how

the independent variables influence the dependent variable individually. This inflates the standard errors of some or all of the regression coefficients.

### **Detecting Multicollinearity**

A statistical technique called the [variance inflation factor](#) (VIF) can detect and measure the amount of collinearity in a multiple regression model. VIF measures how much the variance of the estimated regression coefficients is inflated as compared to when the predictor variables are not linearly related. A VIF of 1 will mean that the variables are not correlated; a VIF between 1 and 5 shows that variables are moderately correlated, and a VIF between 5 and 10 will mean that variables are highly correlated.

When analyzing stocks, you can detect multicollinearity by noting whether the indicators graph the same. For instance, choosing two momentum indicators on a trading chart will generally create trend lines that indicate the same momentum.

## **Reasons for Multicollinearity**

Multicollinearity can exist when two independent variables are highly correlated. It can also happen if an independent variable is computed from other variables in the data set or if two independent variables provide similar and repetitive results.

Again, if you're using the same data to create two or three of the same type of trading indicators, the outcomes will be multicollinear because the data and its manipulation to create the indicators are very similar.

The statistical inferences from a model that contains multicollinearity may not be dependable.

## **Types of Multicollinearity**

### **Perfect Multicollinearity**

Perfect multicollinearity demonstrates a linear relationship that is exact between multiple independent variables. This is usually seen on a chart where the data points fall along the regression line. In technical analysis, it can be seen when you use two indicators that measure the same thing, such as volume. If you overlaid one on top of the other, there would be no difference between them.

### **High Multicollinearity**

High multicollinearity demonstrates a correlation between multiple independent variables, but it is not as tight as in perfect multicollinearity. Not all data points fall on the regression line, but it still signifies data is too tightly correlated to be used.

In technical analysis, indicators with high multicollinearity have very similar outcomes.

### **Structural Multicollinearity**

Structural multicollinearity occurs when you use data to create new features. For instance, if you collected data and then used it to perform other calculations and ran a regression on the results, the outcomes will be correlated because they are derived from each other.

This is the type of multicollinearity seen in investment analysis because the same data is used to create different indicators.

### **Data Based Multicollinearity**

A poorly designed experiment or data collection process, such as using observational data, generally results in data-based multicollinearity, where data is correlated due to the nature of the way it was collected. Some or all of the variables are correlated.

Stock data used to create indicators is generally collected from historical prices and trading volume, so the chances of it being multicollinear due to a poor collection method are small.

## **Multicollinearity in Investing**

For investing, multicollinearity is a common consideration when performing [technical analysis](#) to predict probable future price movements of a security, such as a stock or a [commodity future](#).

Market [analysts](#) want to avoid using technical indicators that are collinear in that they are based on very similar or related inputs; the inputs referred to here are not the data itself but how it was manipulated to achieve the outcome.

Instead, the analysis must be based on markedly different indicators to ensure that the market is analyzed from independent analytical viewpoints. For example, momentum and trend indicators share the same data, but they

will not be perfectly multicollinear or even demonstrate high multicollinearity. These two indicators have different outcomes based on how the data was manipulated.

*Most investors won't worry about the data and techniques behind the indicator calculations—it's enough to understand what multicollinearity is and how it can affect an analysis.*

## **How to Fix Multicollinearity**

One of the most common ways of eliminating the problem of multicollinearity is first to identify collinear independent predictors and then remove one or more of them. Generally, in statistics, a variance inflation factor calculation is run to determine the degree of multicollinearity. An alternative method for fixing multicollinearity is to collect more data under different conditions

## **How Can One Deal With Multicollinearity?**

To reduce the amount of multicollinearity found in a statistical model, one can remove the specific variables identified as the most collinear. You can also try to combine or transform the offending variables to lower their correlation. If that does not work or is unattainable, there are modified regression models that better deal with multicollinearity, such as ridge regression, principal component regression, or partial least squares regression. In stock analysis, the best method is to choose different types of indicators.

## **What Is Multicollinearity in Regression?**

Multicollinearity describes a relationship between variables that causes them to be correlated. Data with multicollinearity poses problems for analysis because they are not independent.

## **How Do You Interpret Multicollinearity Results?**

Data will have high multicollinearity when the variable inflation factor is more than five. If the VIF is between one and five, variables are moderately correlated, and if equal to one, they are not correlated. In technical analysis, the indicators will be generally identical.

## **What Is Perfect Collinearity?**

Perfect collinearity exists when there is an exact 1:1 correspondence between two independent variables in a model. This can be either a correlation of +1.0 or -1.0.

## Why Is Multicollinearity a Problem?

Multicollinearity is a problem because it produces regression model results that are less reliable. This is due to wider confidence intervals (larger [standard errors](#)) that can lower the [statistical significance](#) of regression coefficients. In stock analysis, it can lead to false impressions or assumptions about an investment.

## What Is a Variance Inflation Factor (VIF)?

A variance inflation factor (VIF) is a measure of the amount of multicollinearity in regression analysis. [Multicollinearity](#) exists when there is a correlation between multiple independent variables in a multiple regression model. This can adversely affect the [regression](#) results. Thus, the variance inflation factor can estimate how much the variance of a regression coefficient is inflated due to multicollinearity.

### KEY TAKEAWAYS

- A variance inflation factor (VIF) provides a measure of multicollinearity among the independent variables in a multiple regression model.
- Detecting multicollinearity is important because while multicollinearity does not reduce the explanatory power of the model, it does reduce the statistical significance of the independent variables.
- A large VIF on an independent variable indicates a highly collinear relationship to the other variables that should be considered or adjusted for in the structure of the model and selection of independent variables.

## Understanding a Variance Inflation Factor (VIF)

A variance inflation factor is a tool to help identify the degree of multicollinearity. Multiple regression is used when a person wants to test the effect of multiple variables on a particular outcome. The dependent variable is the outcome that is being acted upon by the independent variables—the

inputs into the model. Multicollinearity exists when there is a linear relationship, or correlation, between one or more of the independent variables or inputs.

### **The Problem of Multicollinearity**

Multicollinearity creates a problem in the multiple regression model because the inputs are all influencing each other. Therefore, they are not actually independent, and it is difficult to test how much the combination of the independent variables affects the dependent variable, or outcome, within the regression model.

While multicollinearity does not reduce a model's overall predictive power, it can produce estimates of the regression coefficients that are not statistically significant. In a sense, it can be thought of as a kind of double-counting in the model.