

## 1. What is Variance Inflation factor?

The Variance Inflation Factor (VIF) is a measure used to assess the severity of multi-collinearity in a regression analysis. Multi-collinearity occurs when independent variables in a regression model are highly correlated with each other, which can lead to unstable and unreliable estimates of the regression coefficients. The VIF quantifies how much the variance of the estimated coefficients is inflated due to multi-collinearity.

**Here's how VIF works:**

**Step 1:** For each independent variable in your regression model, you fit a separate regression model where that variable is the dependent variable, and all other independent variables are used as predictors.

**Step 2:** Calculate the VIF for each independent variable using the formula:

$$\text{VIF} = 1 / (1 - R^2)$$

$R^2$  is the coefficient of determination from the regression model in Step 1. It represents the proportion of the variance in the variable explained by the other independent variables.

**Interpretation:** The VIF quantifies how much the variance of the estimated coefficient for a particular variable is inflated due to multi-collinearity. Specifically, a VIF of 1 indicates no multi-collinearity (i.e., the variable is not correlated with any other predictors), and higher VIF values indicate more severe multi-collinearity.

Common guidelines for interpreting VIF values include:

VIF = 1: No multi-collinearity (no correlation with other predictors).

VIF < 5: Mild multi-collinearity (usually considered acceptable).

VIF  $\geq$  5 and < 10: Moderate multi-collinearity (may warrant further investigation or action).

VIF  $\geq$  10: Severe multi-collinearity (often considered problematic).

When you encounter variables with high VIF values, you may consider the following actions:

**Remove one or more correlated variables:** If two or more variables have high VIF values, you may choose to remove one of them to mitigate multi-collinearity.

**Combine correlated variables:** You can create new variables through feature engineering that capture the combined information of correlated variables, reducing multi-collinearity.

**Regularization techniques:** Consider using regression techniques like Ridge regression or Lasso regression, which automatically address multi-collinearity by adding penalty terms to the regression equation.

VIF is a valuable tool for identifying and quantifying multi-collinearity, helping you make informed decisions about how to handle it in your regression analysis.