* M**ulticollinearity** (also **collinearity**) is a phenomenon in which one feature variable in a regression model is highly linearly correlated with another feature variable.
* A **collinearity** is a special case when two or more variables are exactly correlated.
* In the presence of **multicollinearity**, regression estimates are unstable and have high standard errors.

|  |  |
| --- | --- |
| **collinearity** | **multicollinearity** |
| * **Collinearity** is a linear association **between** two predictors. | * **Multicollinearity** is a situation where two or more predictors are highly linearly related. |

**is Collinearity bad?**

* The coefficients become very sensitive to small changes in the model.
* **Multicollinearity** reduces the precision of the estimate coefficients, which weakens the statistical power of your regression model. You might not be able to trust the p-values to identify independent variables that are statistically significant.

**Causes of collinearity :**

* It is **caused** by an inaccurate use of dummy variables.
* It is **caused** by the inclusion of a variable which is computed from other variables in the data set.
* **Multicollinearity** can also result from the repetition of the same kind of variable. Generally occurs when the variables are highly correlated to each other.

**How is Vif calculated?**

* The Variance Inflation Factor (**VIF**) is a measure of colinearity among predictor variables within a multiple regression.
* It is **calculated** by taking the the ratio of the variance of all a given model's betas divide by the variane of a single beta if it were fit alone.

**How to Deal With Multicollinearity?**

* Remove highly correlated predictors from the model. ...
* Use Partial Least Squares Regression (PLS) or Principal Components Analysis, regression methods that cut the number of predictors to a smaller set of uncorrelated components.

**https://www.analyticsvidhya.com/blog/2020/03/what-is-multicollinearity/**