MULTI – COLLINEARITY

What is multi-collinearity ?

Multi-collinearity occurs when two or more independent variables in a regression model are highly correlated, indicating a strong linear relationship among the predictor variables.

Effects of multi-collinearity.

1. **Uncertainty in coefficient estimates or unstable variance:** Small changes (adding/removing rows/columns) in the data results in change of coefficients.
2. **Increased standard error:** Reduces the accuracy of the estimates and increases the chances of detection.
3. **Decreased statistical significance:** Due to increased standard error, t-statistic declines which negatively impacts the capability of detecting statistical significance in coefficient leading to type-II error.
4. **Reducing coefficient & p-value:** The importance of the correlated explanatory variable is masked due to collinearity.
5. **Overfitting:** Leads to overfitting as is indicated by the high variance problem.

Handling multi-Collinearity.

Multi-collinearity can be handled with the following two methods.

1. **Introduce penalization or remove highly correlated variables:** Use lasso and ridge regression to eliminate variables which provide information which is redundant. This can also be achieved by observing the VIF.
2. **Combine highly correlated variables:** Since the collinear variables contain redundant information, combining them into a single variable using methods such as PCA to generate independent variables.