Multiple Linear Regression Analysis

Benefits:

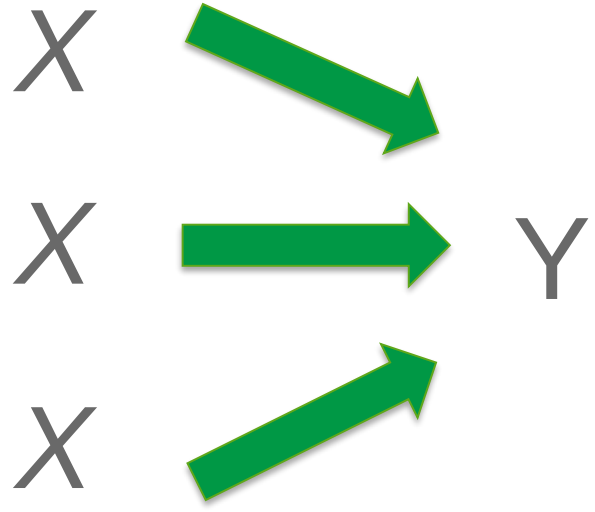
- Expand regression capabilities to handle multiple input variables
- Ability to forecast
- Enhance existing analytics platform investments
- Increase analytical access to a wider community

Expanding Linear Regression

Simple Linear Regression

One input variable and one output variable

Multiple Linear Regression

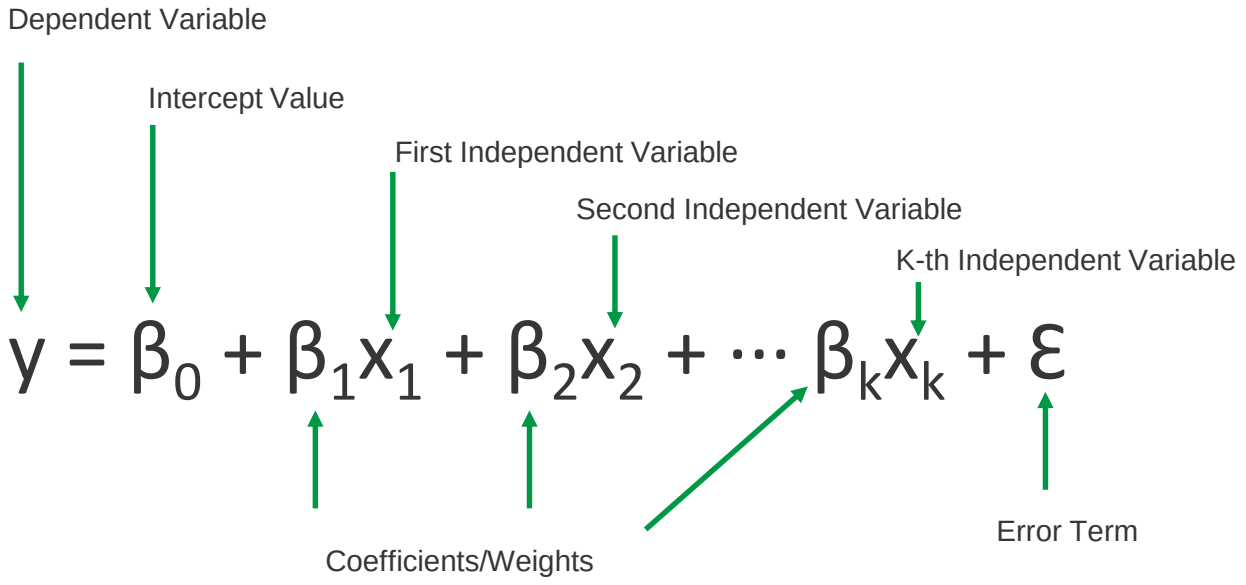


multiple input variable and one output variable

Multiple Linear Regression Analysis

Formula

The following formula is utilized to power multiple linear regression analysis



Multiple Linear Regression Analysis

Correlation

Existence of a relationship between two or more variables; measured numerically as the correlation coefficient

Closer to 1.0 or -1.0 = Stronger

Closer to 0 = Weaker

R-Squared

Proportion of the variance in the dependent variable that is predictable from the independent variable; often derived by squaring correlation coefficient

R-Squared Adjusted

Modified version of R-Squared that increases only if the variables improves the model more than would be expected by chance

Closer to 1.0 = Stronger

Closer to 0 = Weaker

Multiple Linear Regression Analysis

Feature Selection

Process of selecting a subset of relevant features or variables for use in model construction

Benefits:

- Reduces Overfitting by eliminating excess variables
- Creates smaller and more efficient models
- Highlights and eliminates costly low performing variables

Common Automatic Feature Selection Techniques

Forward Selection - start with a single variable and add potential variables one by one until there is nothing useful left to add

Backward Selection - start with all potential variables and remove variables one by one until everything remaining is useful

Stepwise Selection – combine forward and backward selection by starting with one variable and then adding variables; after each addition, check if any variable can be can be removed

Multiple Linear Regression Analysis

Watch-outs

To ensure multiple linear regression analysis is appropriate, these watch-outs need to be accounted for

1. *Ensure a linear relationship exists between the variables you are looking the analyze*
2. *Address potential overfitting issues that may occur*
3. *Investigate and mitigate the presence of multicollinearity within predictor variables*
4. *Verify all of the assumptions required for utilizing linear modeling including, normality and heteroscedasticity in the error terms*