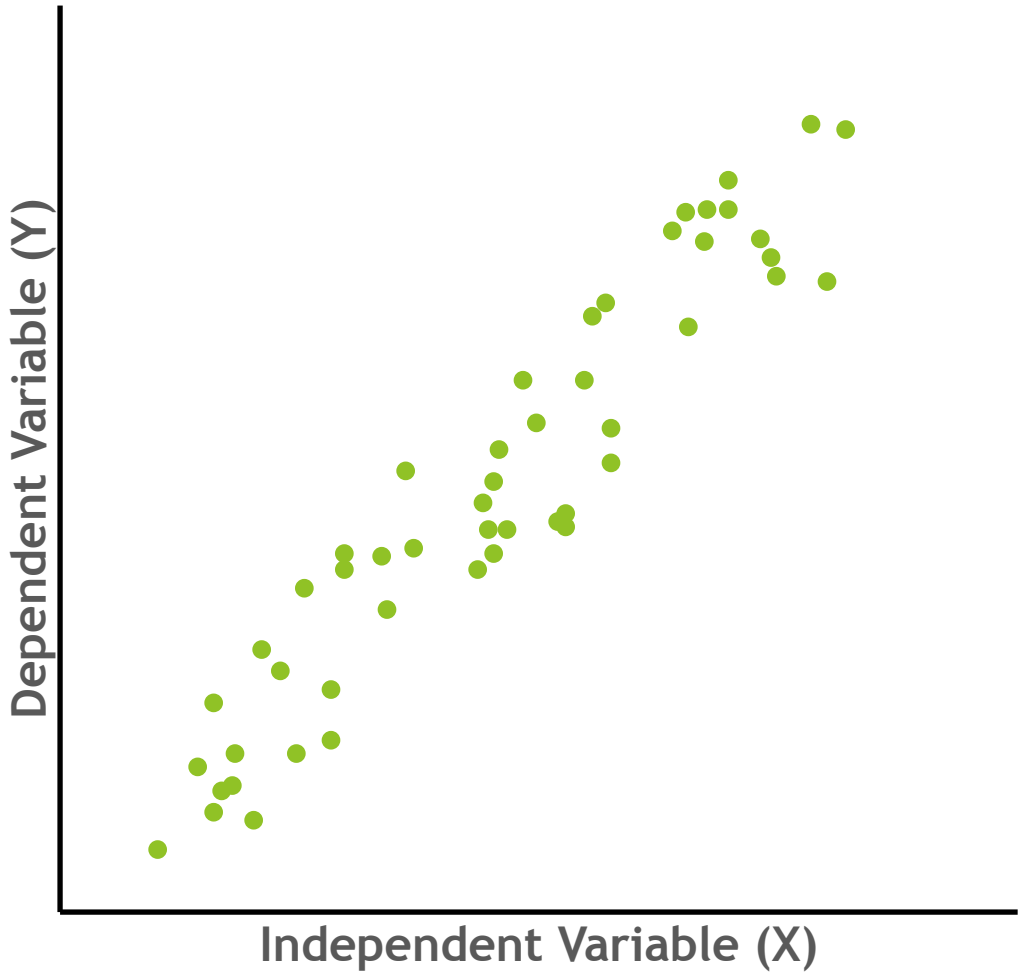
Linear Regression

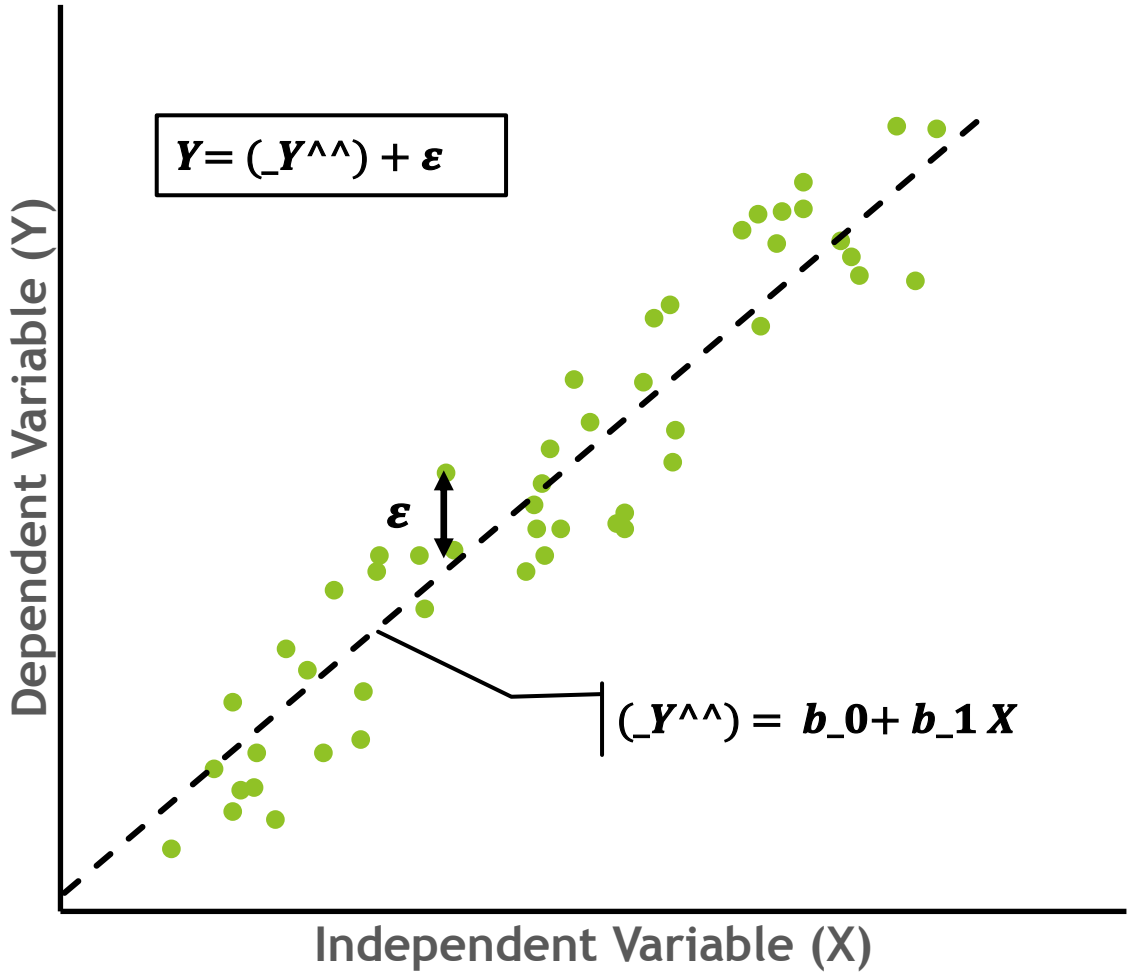
Regression

- Regression explains the variation in a dependent variable using the independent variables.
- The output of a regression is a function that predicts the dependent variable based upon values of the independent variables.



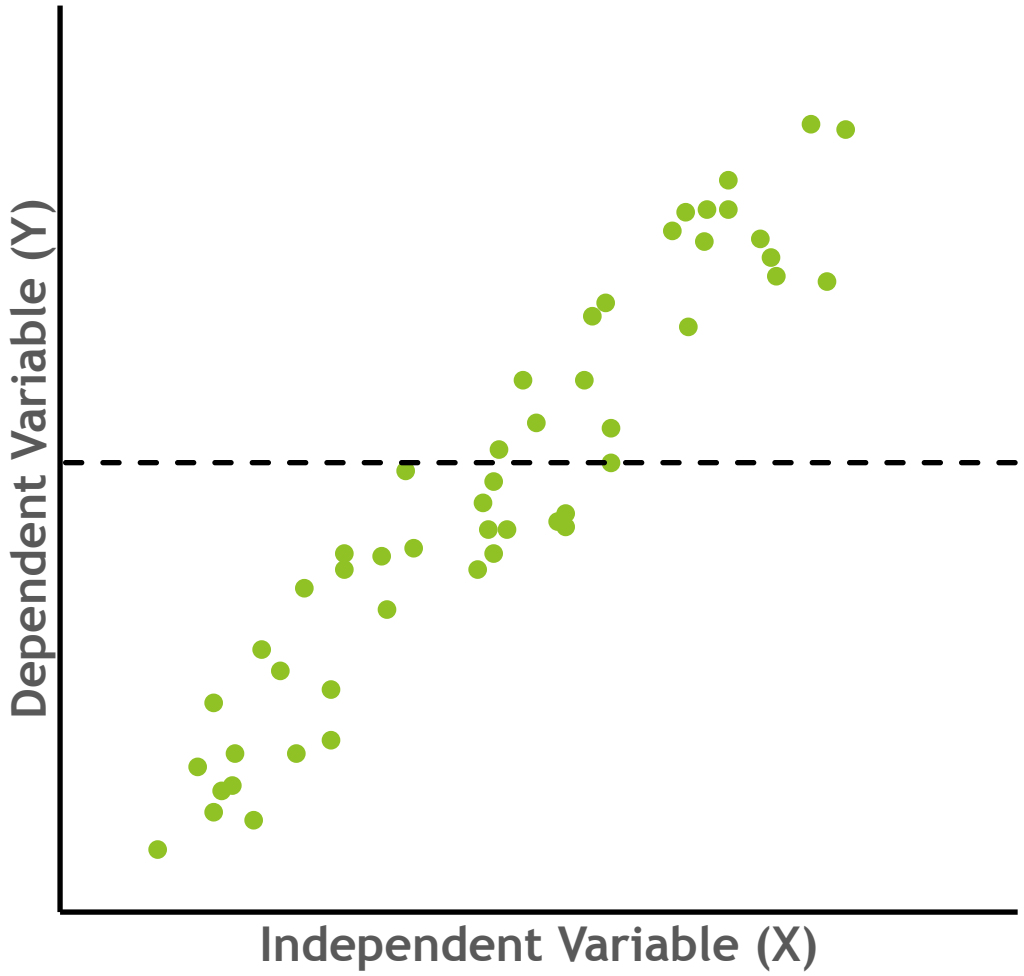
Linear Regression & SSE

- Linear regression fits a straight line to the data.
- A least squares regression selects the line with the lowest total sum of squared prediction errors or SSE.



Linear Regression & SSR

- The Sum of Squares Regression (SSR) is the sum of the squared differences between the prediction for each observation and the population mean.



SST & R^2

- Sum of Squares Total (SST)
- The proportion of total variation (SST) that is explained by the regression (SSR) is known as the Coefficient of Determination, and is often referred to as R^2.

$$SST = SSR + SSE$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Degrees of Freedom & Standard Error

- **Degrees of Freedom (df) in a statistical calculation represent how many values involved in a calculation have the freedom to vary**

$$df = n - k$$

- **The Standard Error of a regression is a measure of its variability**

$$SE = \sqrt{\frac{SSE}{n-k}}$$