Regression Analysis – Multiple Regression Model
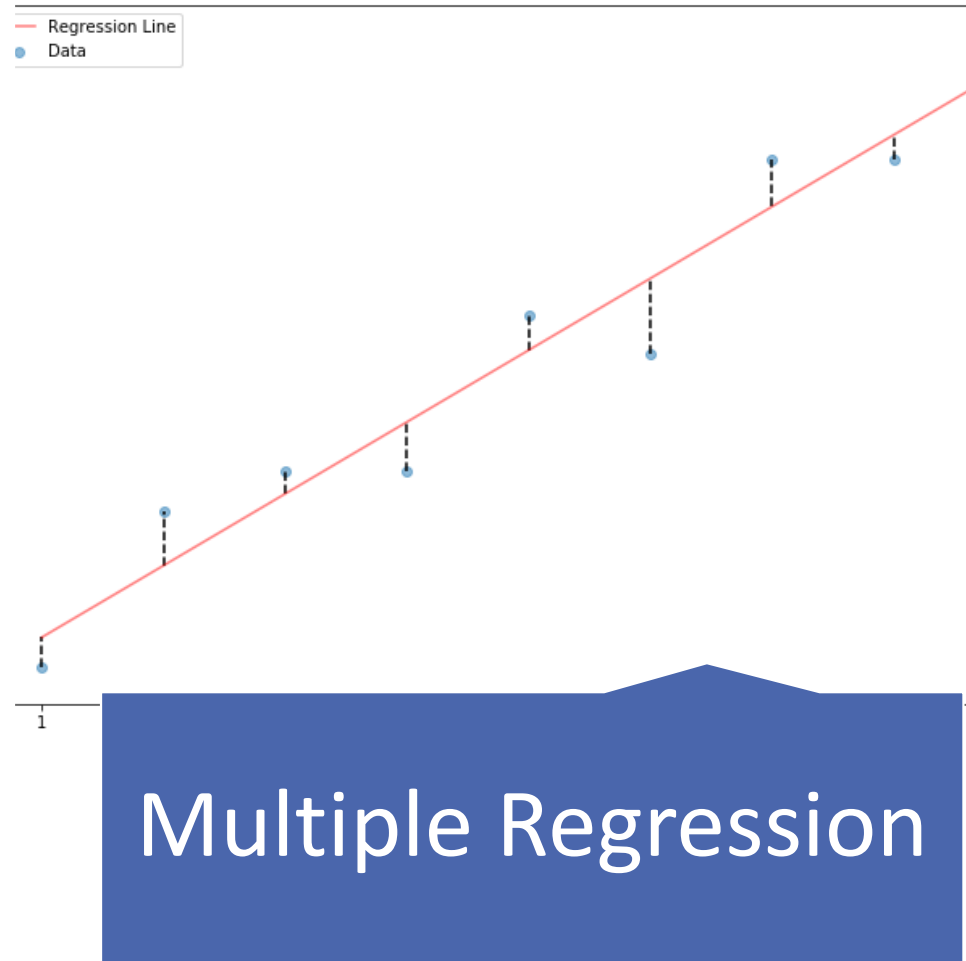

Multiple Regression

Multiple Regression

The purpose of Multiple Linear Regression is to explain (the variation in) a dependent variable by (the variation in) two or many independent variables.

- dependent Variable: Movie Revenue (Y)
- independent Variables: Movie Budget, Runtime, Popularity, average Rating, etc… (X_1, X_2, X_3)

$\rightarrow$ Setting up a multiple linear model:

$$Y_i = a + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_k X_{ki} + \varepsilon_i$$

Revenue   Budget   Runtime   Rating

Multiple Regression – Properties

Properties

- hard to visualize → higher dimensional space

- OLS Regression → minimize MSE / SSE!

- Intercept (a) is the value of the dependant variable when the independent variables are equal to zero

- Each Slope Coefficient (b) is the estimated change in the dependent variable for a one-unit change in that independent variable – holding the other independent variables constant → partial slope coefficients

- Conduct a t-test on each slope coefficient to determine the significance of each independent variable

Multiple Regression – Implications

Implications

- A multiple linear regression model captures correlations between independent variables.

- By adding an additional independent variable:

  - The slope coefficients of other independent variables (can) change

  - The p-values of other independent variables (can) change

  - Significance of other independent variables (can) change → Remove

- Omitting important independent variable(s) can distort the model.

- Highly correlated independent variables can lead to Multicollinearity and requires model re-specification (see next section).