Start-Tech Academy

Bivariate Analysis

Bivariate analysis is the simultaneous analysis of two variables (attributes). It explores the concept of relationship between two variables, whether there exists an association and the strength of this association, or whether there are differences between two variables and the significance of these differences.

Creating new Variables

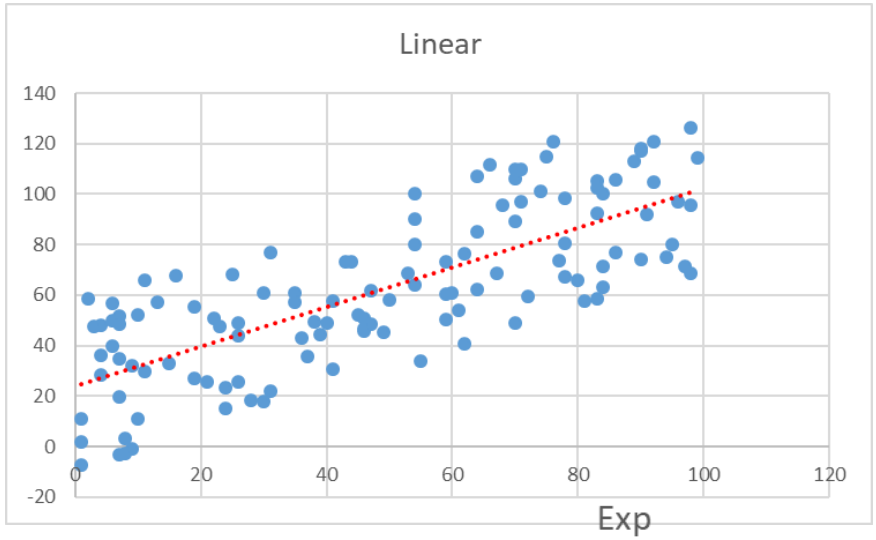
Scatter Plot

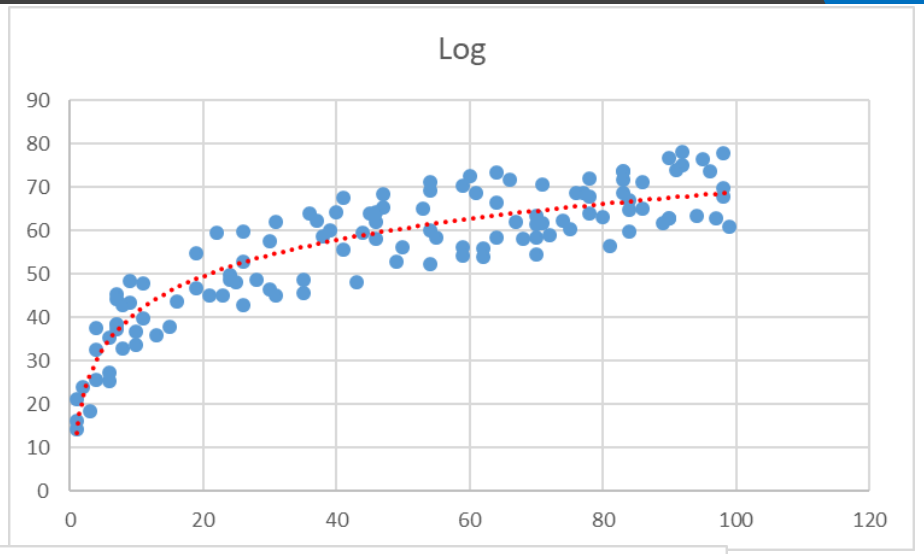
- Scatter indicates the type (linear or non-linear) and strength of the relationship between two variables
- We will use Scatter plot to transform variables

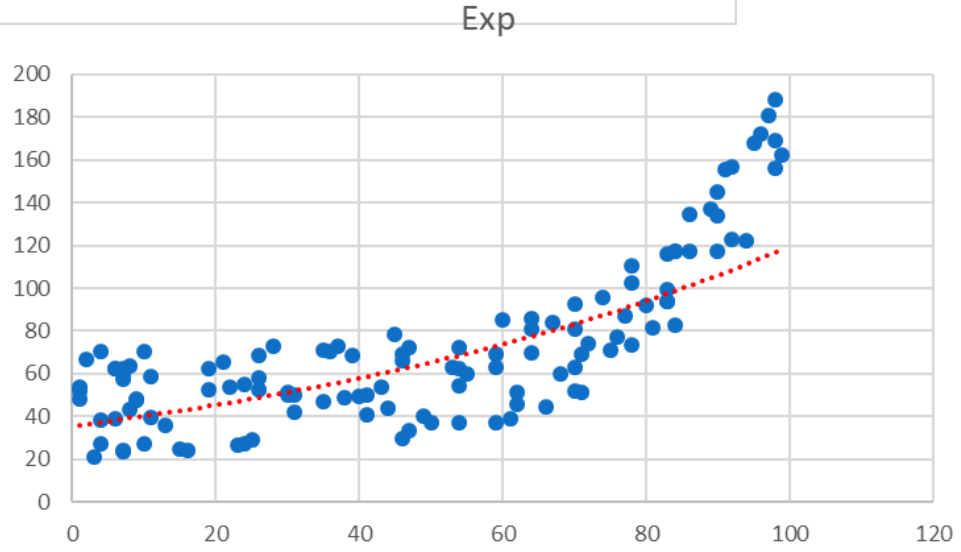
Correlation

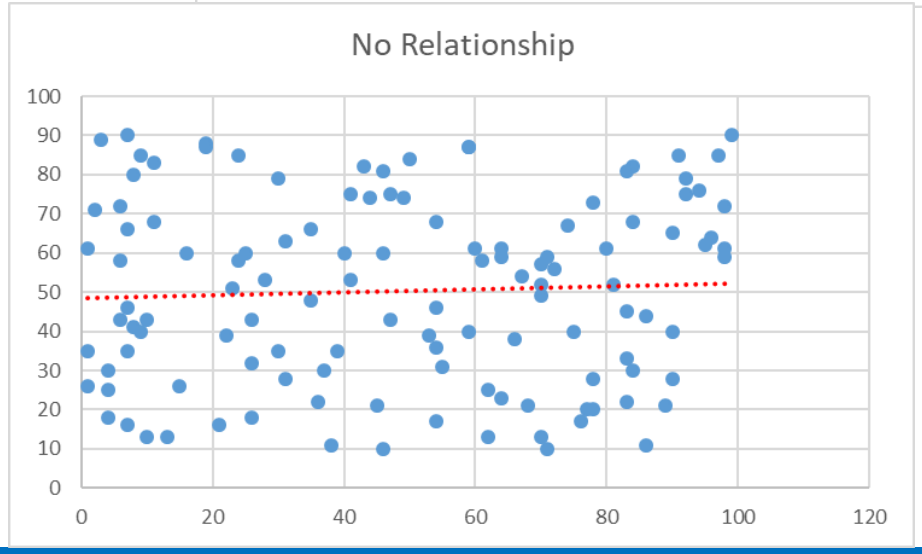
- Linear correlation quantifies the strength of a linear relationship between two numerical variables.
- When there is no correlation between two variables, there is no tendency for the values of one quantity to increase or decrease with the values of the second quantity.
- Correlation is used to drop Non Usable variables

Scatter plots









Variable Transformation

Transform your existing variable to extract more information out of them

Creating new Variables

Identify

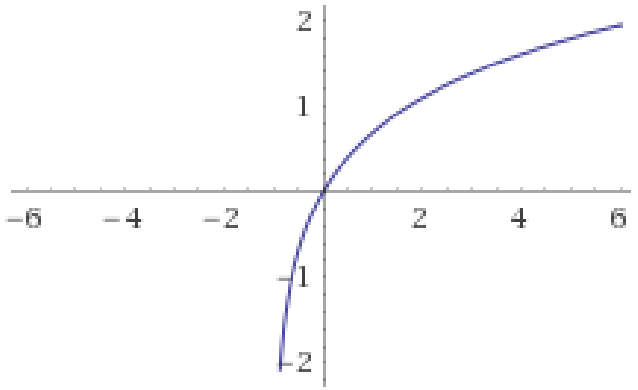
- Using your business knowledge and bivariate analysis to modify variable

Methods

- Use Mean/Median of variables conveying similar type of information
- Create ratio variable which are more relevant to business
- Transform variable by taking log, exponential, roots etc.

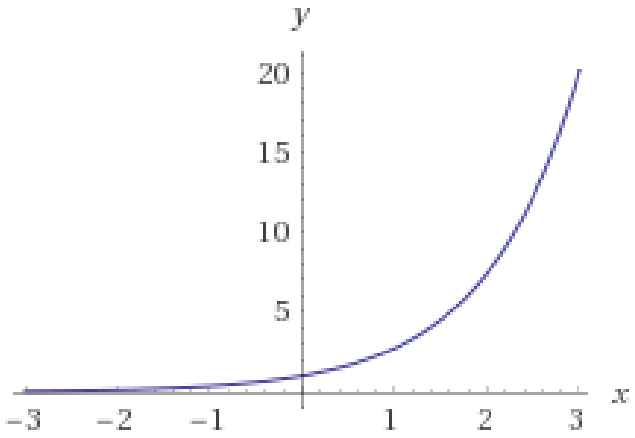
Transformation

If



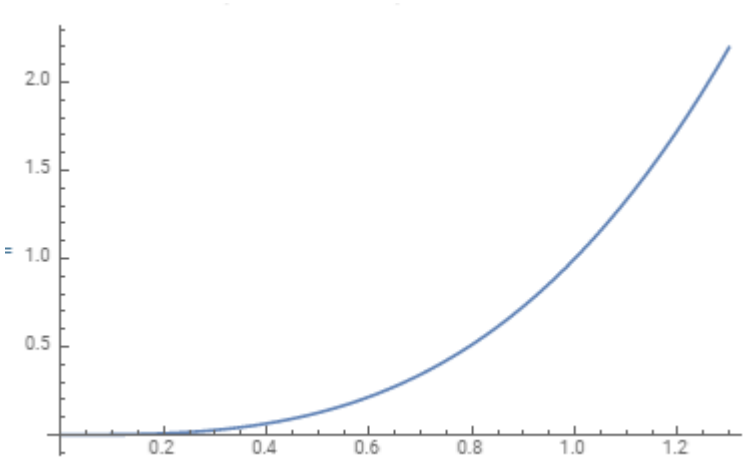
Take log(1+x) instead of x

If



Take e^x instead of x

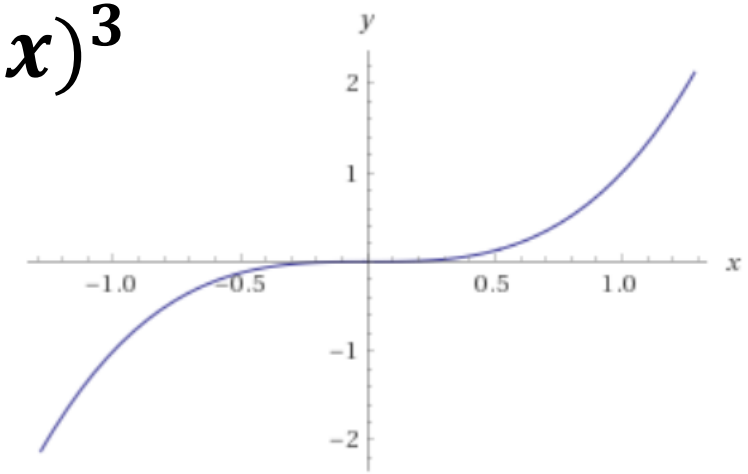
Transformation

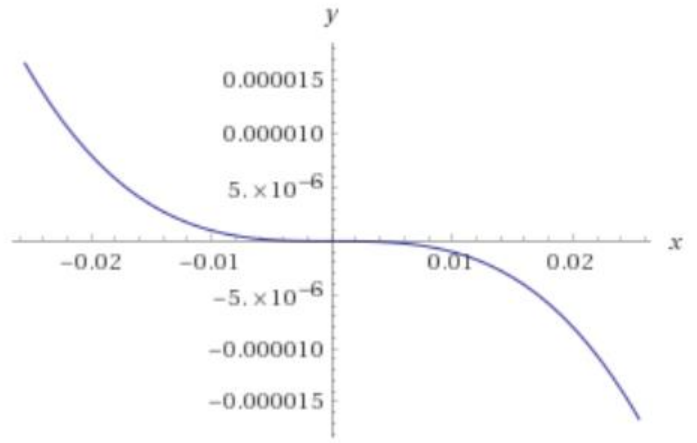
If 

Take x^2 or x^n instead of x

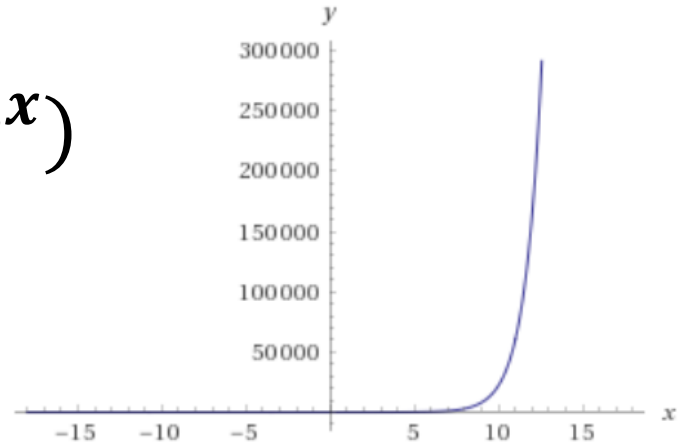
Effect of Putting Negative Sign

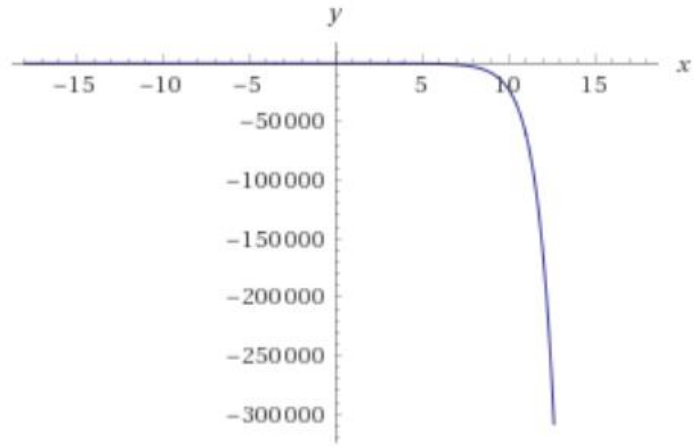
x^3 *vs* $(-x)^3$





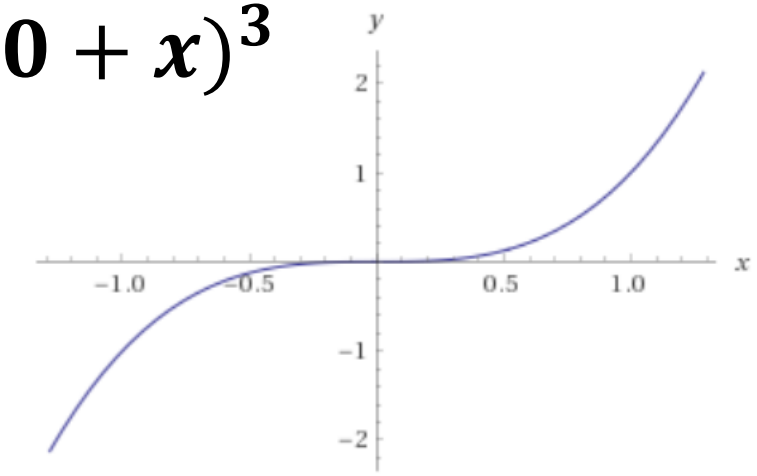
e^x *vs* $(-e^x)$

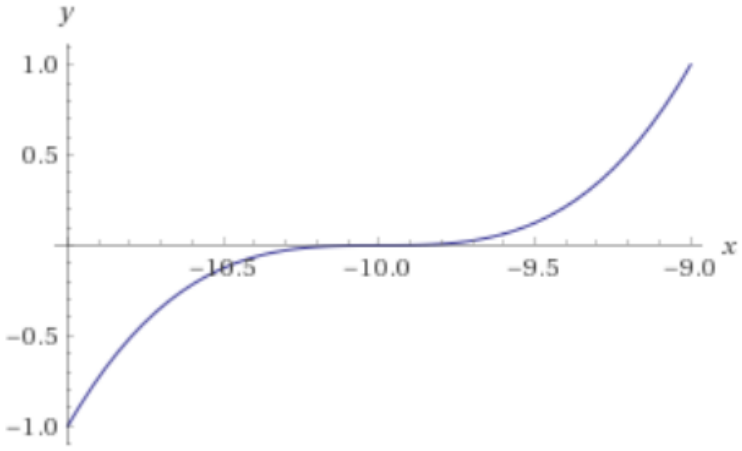
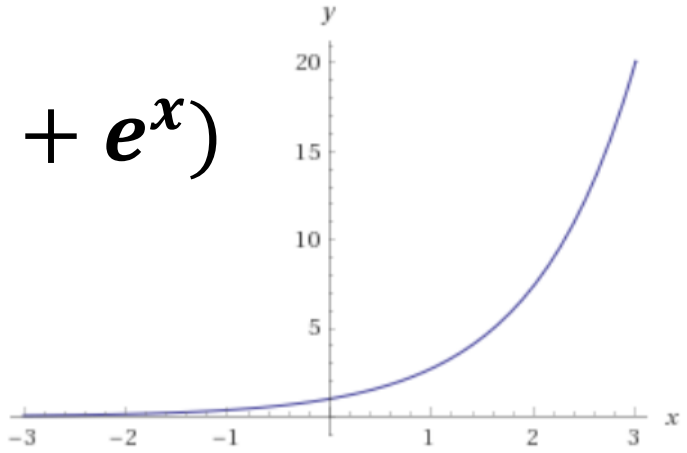
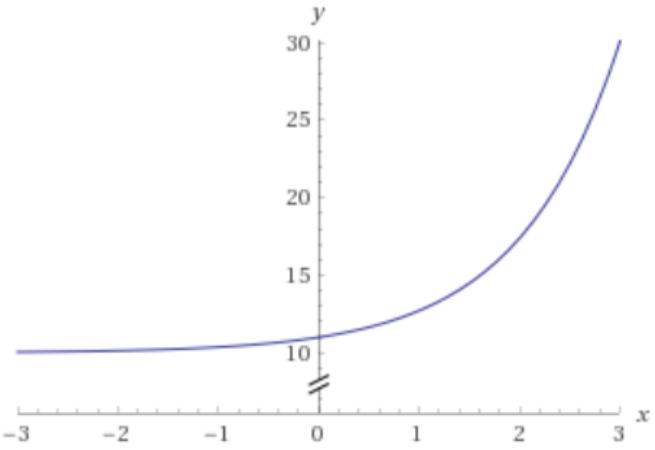




Effect of Adding a constant

$$x^3 \; vs \; (10 + x)^3$$





$$e^x \; vs \; (10 + e^x)$$