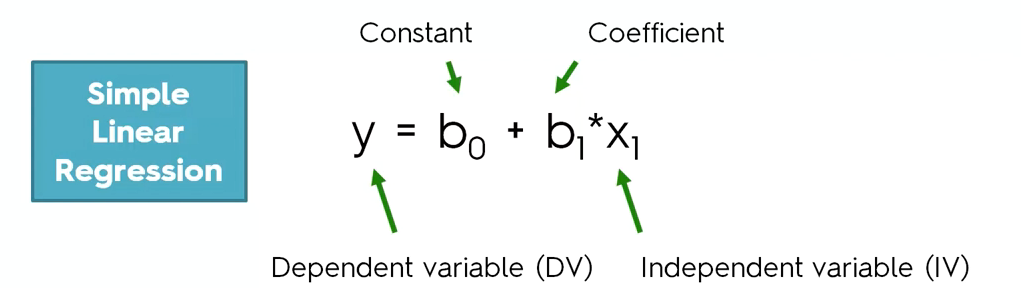
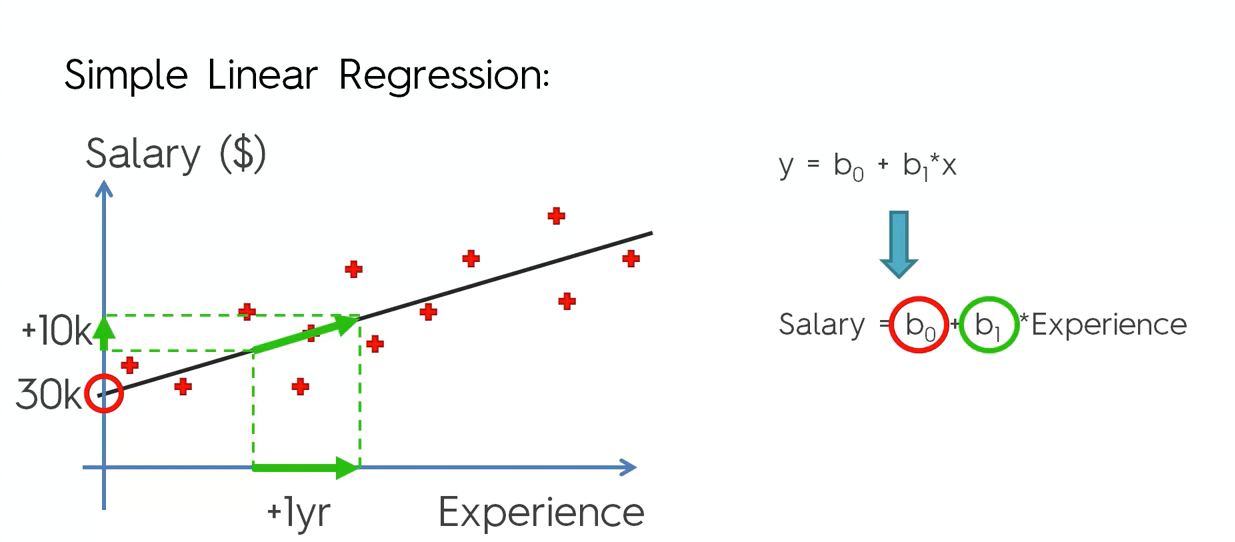
Regression

**Simple Linear Regression:**

It is like a formula of a slope

Eg scenario: how does a **salary** (Dependent) changes with **years of experience** (Independent Variable)

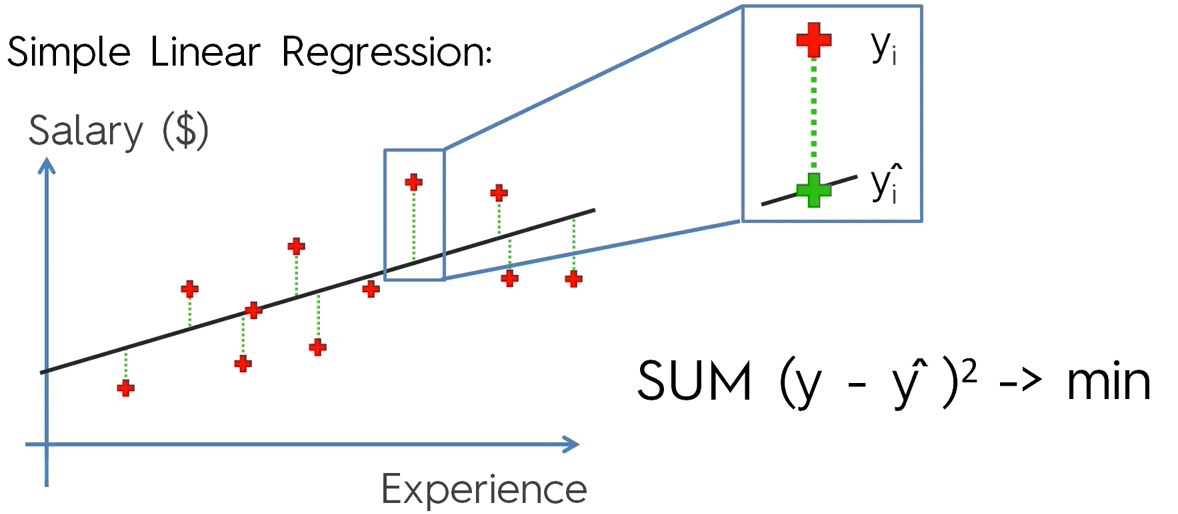
Coefficient🡪 Unit change in independent Variable (like multiplier🡪connection between x1 and y)



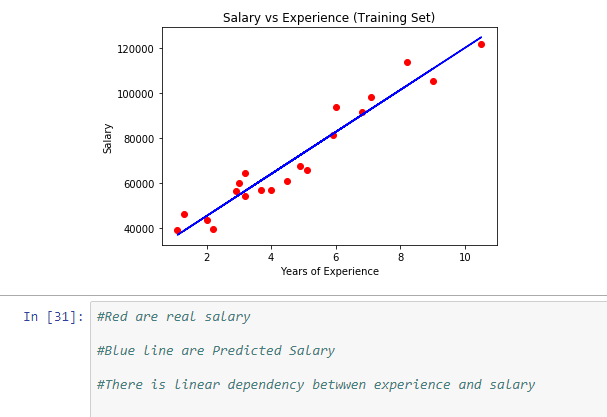
b0🡪Constant (30,000$)

b1🡪Coefficient (+10,000$ for 1 year)

1. How to find the best fitting line(line in the graph)



**Example**



Correlation🡪The correlation coefficient is a **statistical measure** that calculates the strength of the relationship between the relative movements of the two variables

It measures the strength (qualitatively) and direction of the **linear** relationship between two or more variables. The Pearson correlation coefficient measures the strength of the **linear** association between two variables

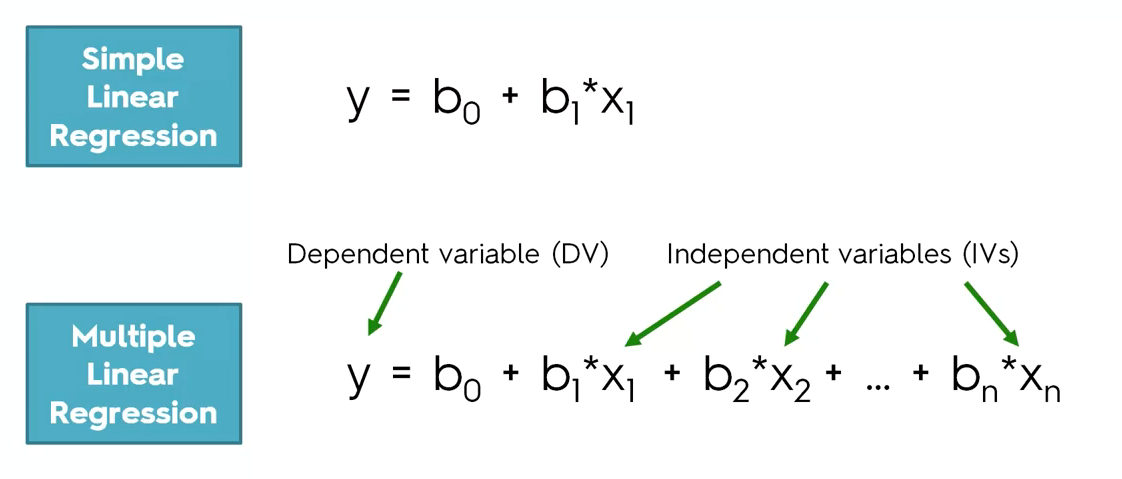
**DIFFERENCE BETWEEN SIMPLE AND LINEAR REGRESSION:**

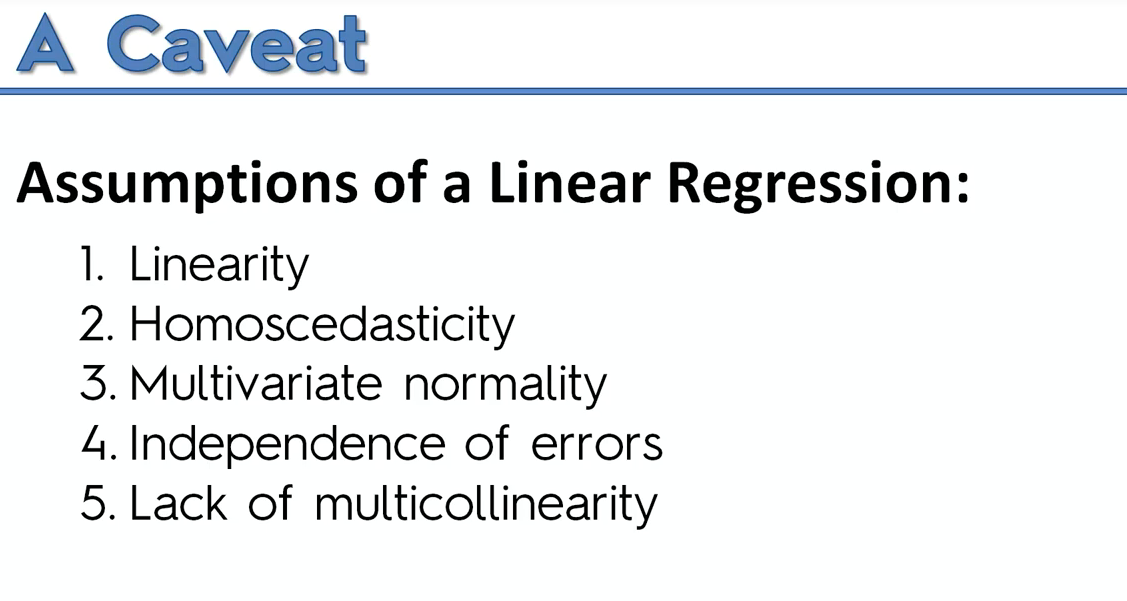
**Simple linear regression:**a single independent variable is used to predict the value of a dependent variable.

Equation: **y=A+BX**

**Multiple linear regression:** two or more independent variables are used to predict the value of a dependent variable. The difference between the two is the number of independent variables.

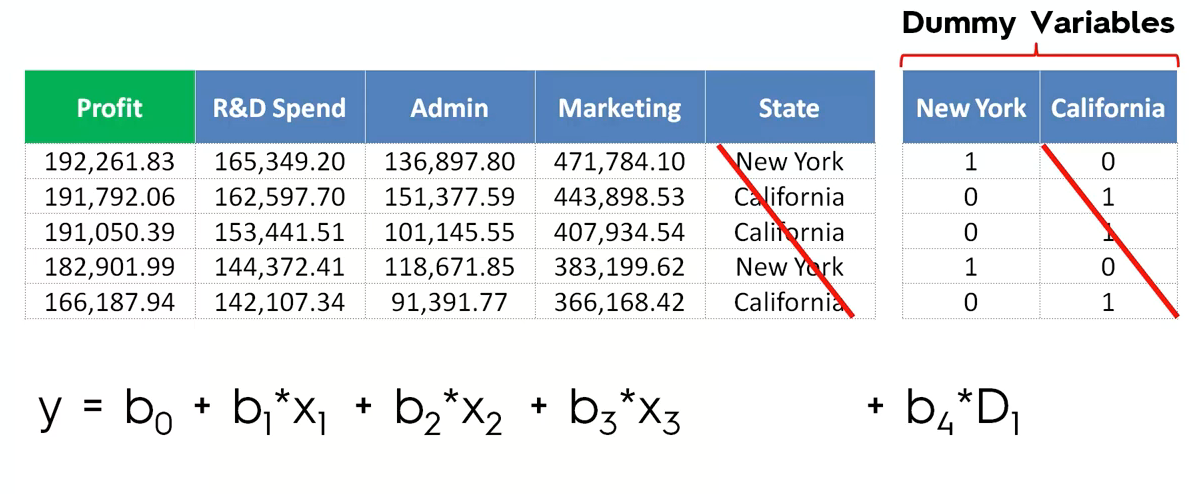
Equation: **y=A+BX1+CX2+DX3**





**Check the above factors before jumping into building linear regression**

**Multiple Linear Regression:**



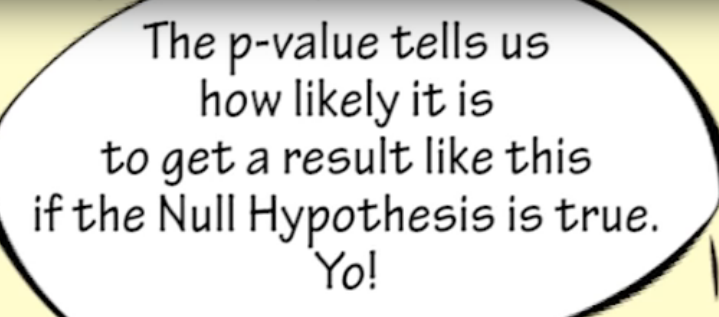
Green color(Profit)🡪dependent variable(y)

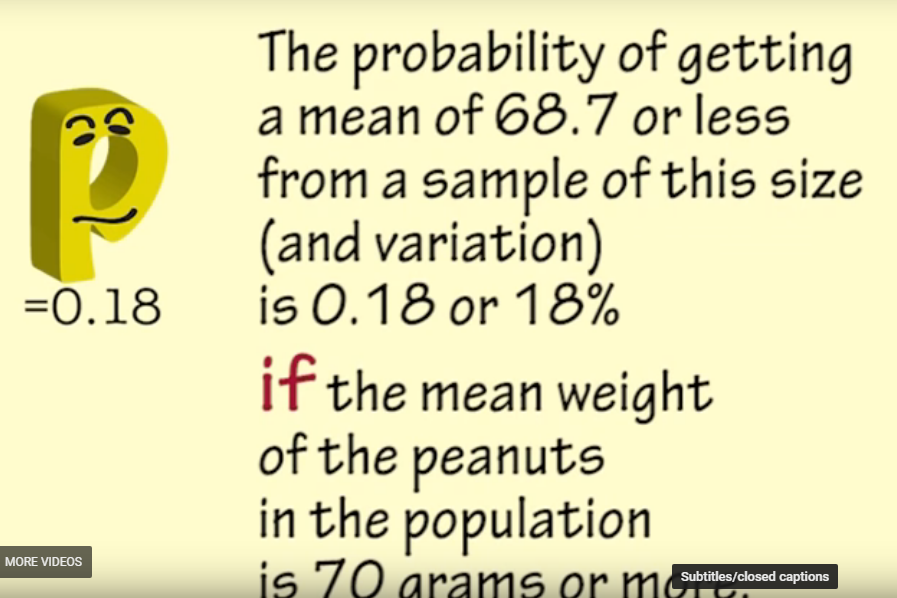
Blue color tabular col🡪independent variable(b1—Coefficient,x1🡪variable)

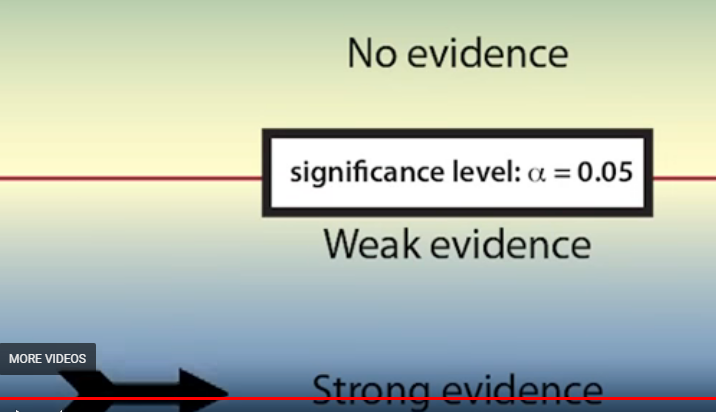
B0🡪 constant

State🡪categorical data, hence we are using dummy variable.

**P-Value:**

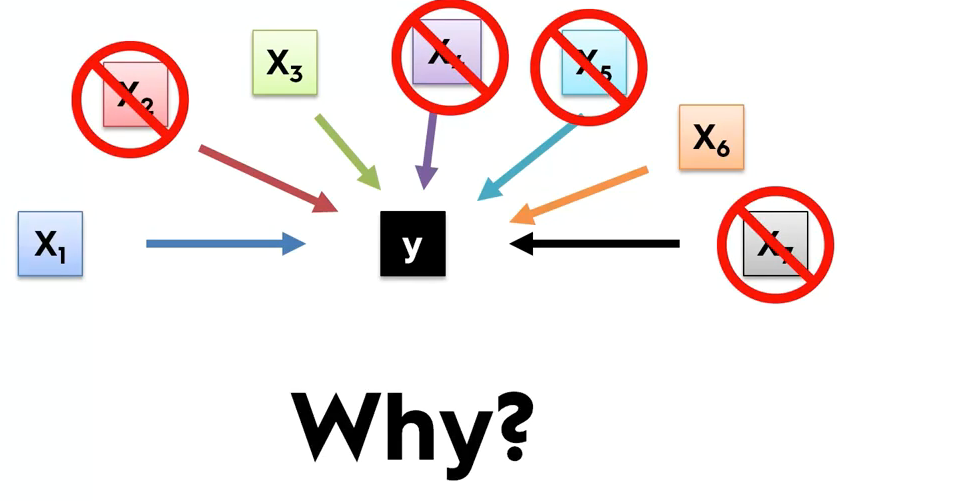






**Building a model step by step:**

We need to remove some independent variable,

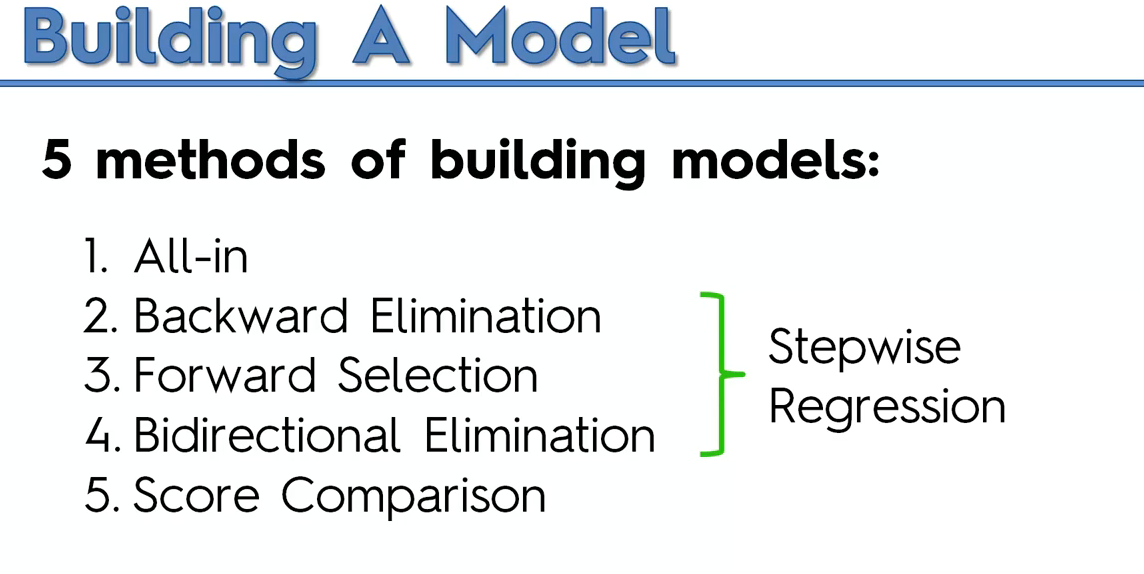


**Reason 1:**

Garbage in –Garbage out🡪 if you through in lot of stuff, then It won’t be good and accurate and it will be a garbage model.

**Reason 2:**

It may confuse the system, hence we should select the independent variable carefully.



Refer the “**Building a model**” PDF to know more

Backward Elimination is the best and efficient.

Score comparison consumes so much of the resource and the complexity grows exponentially .