Correlation, Regression

Presented By
Aditya Joshi
Assistant Professor
Graphic Era Deemed to be University

Correlation

- It is used to test relationship between quantitative variable or categorical variable. It measures how things are related. The study how variables are correlated is called correlation analysis.

High correlation:

your calorie intake & your weight.

Your eye color and your relatives eye color.

The amount of time you study and your GP Score.

Low correlation:

A dog name and type of biscuit they prefer.

The cost of carwash and how long it takes to buy and soda inside the station.

Relations are useful because if you find what relationship variables have, you can make predictions about future behaviour.

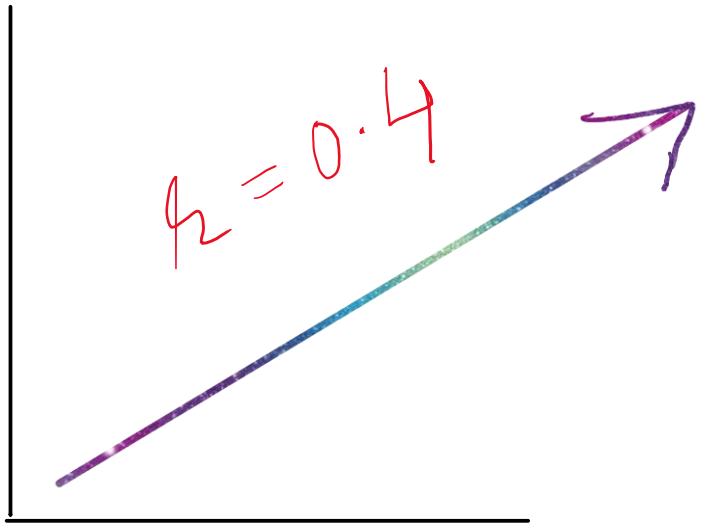
Correlation coefficient: it is used to measure how strong relationship is between the variables. It is a way to put a value to the relationship.

It has value between -1 & 1

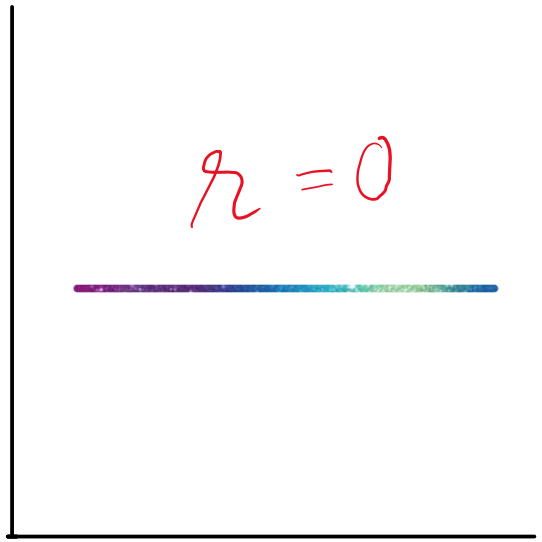
0 means no relationship b/w variable at all.

-1 means there is a perfect negative correlation.

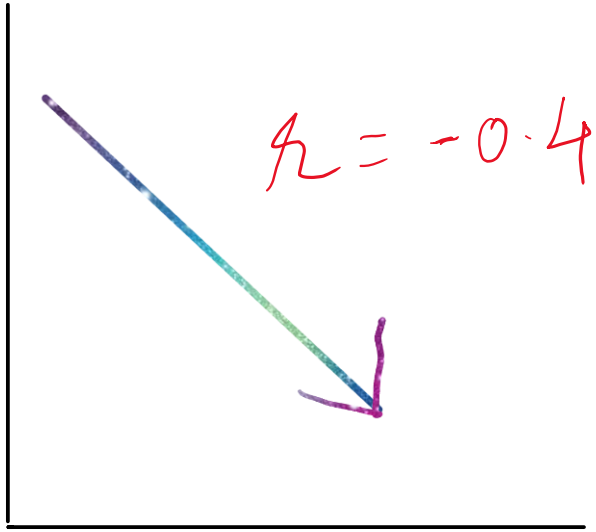
1 means there is a perfect positive correlation.



Positive



No Correlation



Negative

Pearson correlation formula:

$$r=\frac{n(\sum xy)-(\sum x)(\sum y)}{\sqrt{[n\sum x^2-(\sum x)^2][n\sum y^2-(\sum y)^2}}$$

$n=6$

$\sum xy = 20485$

$\sum x = 247$

$\sum y = 486$

$\sum x^2 = 11409$

$(\sum x)^2 = 61009$

$\sum y^2 = 40022$

$(\sum y)^2 = 236196$

Find the value of correlation coefficient from the following table

Subject	Age x	Glucose level y
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

Regression Modeling

- Regression analysis is a way to find trends in data. For example, you might guess that there's a connection between how much you eat and how much you weigh; regression analysis can help you quantify that.
- Regression analysis will provide you with an equation for a graph so that you can make predictions about your data.

For example, if you've been putting on weight over the last few years, it can predict how much you'll weigh in ten years time if you continue to put on weight at the same rate.

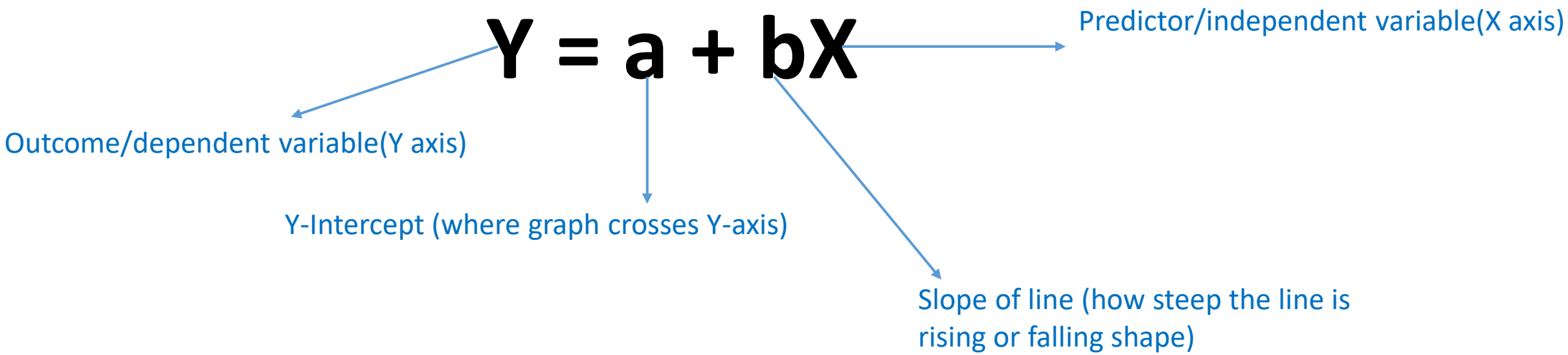
- It will also give you p-value and correlation coefficient to tell you how accurate your model is.
- Elementary stats use very basic techniques like making scatter plot and performing linear regression and multiple regression also

A Simple Linear Regression:

Is where the relationships between your variables can be described with straight line non linear regression produces curved lines

Simple linear regression plots one independent variable X against another dependent variable Y

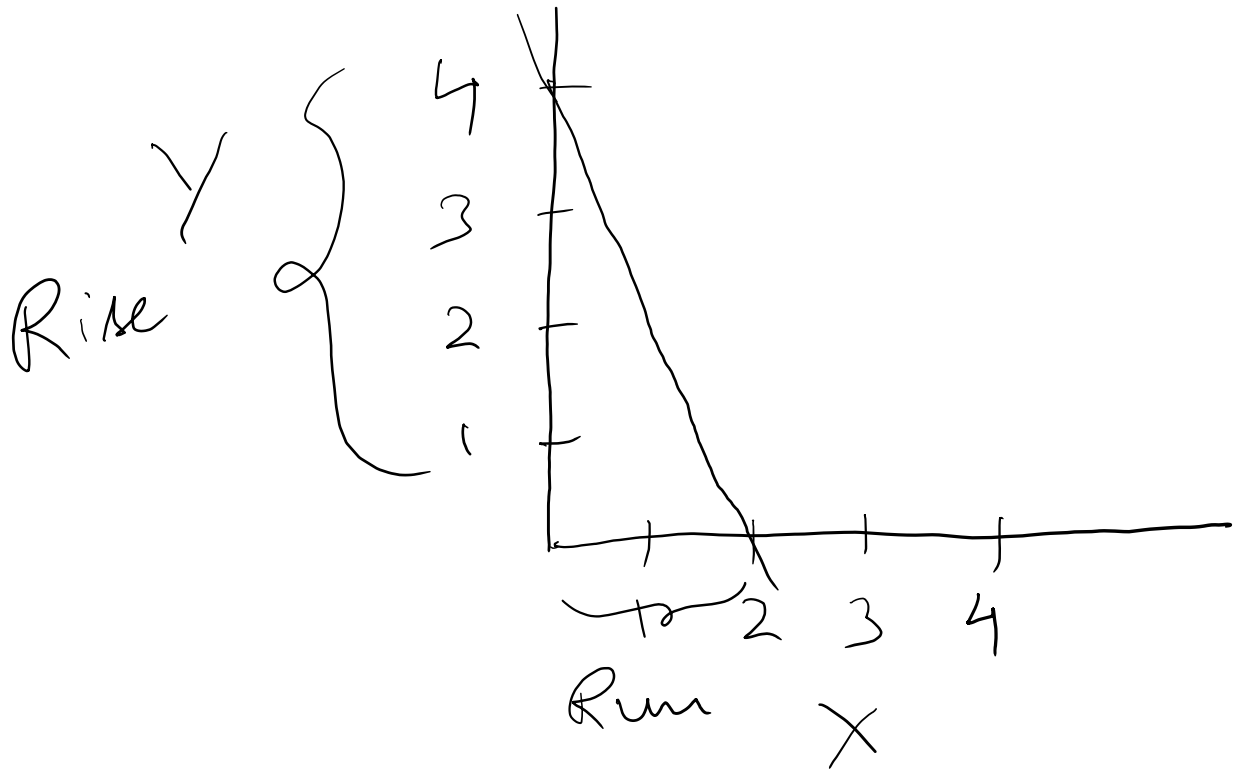
It is given by a linear equation:



Ex → $Y = 4 + 2X$

4 → Y-intercept

2 → Slope $= \frac{\text{Rise}}{\text{Run}} = \frac{4}{2} = 2$



Multiple Linear Regression Analysis:

It is used to see if there is a statistically significant relationship between set of variables.

It is used to find the trends in the set of data

It is almost same as simple linear regression the difference is number of predictor X variable

Simple linear regression uses single X variable for each dependent Y variable(X_1, Y_1)

Multiple linear regression uses multiple X variables (X_1,X_2,X_3, Y_1)

$$\mathbf{Y_1=(b_0+b_1X_1)+(b_0+b_1X_2)+(b_0+b_1X_3)\ldots\ldots\ldots\ldots(b_0+b_1X_n)}$$