ambitious baba .com
INDIA'S LARGEST TEST PREP PLATFORM

CAIIB PAPER-1

Module-A Unit-4

Advanced Bank Management (ABM)

CAIIB Paper 1 (ABM) Module A Unit 4: Correlation and Regression

Introduction

Correlation Analysis

- **Correlation analysis is applied in quantifying the association between two continuous variables**, for example, an dependent and independent variable or among two independent variables.

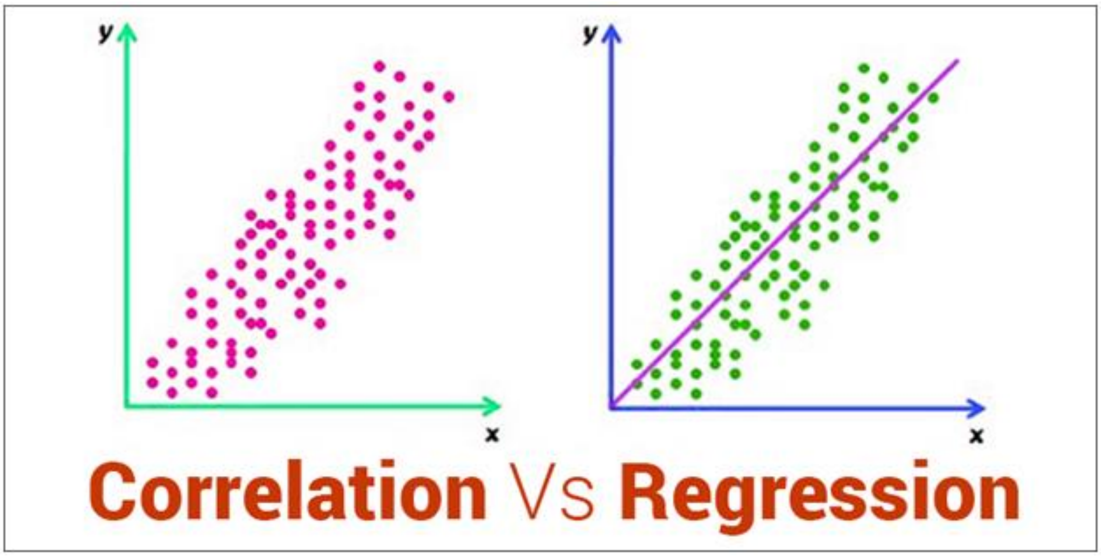
Regression Analysis

- **Regression analysis refers to assessing the relationship between the outcome variable and one or more variables**. The outcome variable is known as the dependent or response variable and the risk elements, and cofounders are known as predictors or independent variables.
- **The dependent variable is shown by "y" and independent variables are shown by "x" in regression analysis.**

Linear Regression

- Linear regression is a **linear approach to modelling the relationship between the scalar components and one or more independent variables**. If the regression has one independent variable, then it is known as a simple linear regression. If it has more than one independent variables, then it is known as multiple linear regression.
- Linear regression only focuses on the conditional probability distribution of the given values rather than the joint probability distribution. In general, all the real world regressions models involve multiple predictors. So, the term linear regression often describes multivariate linear regression.

Correlation and Regression Differences



There are some differences between Correlation and regression.

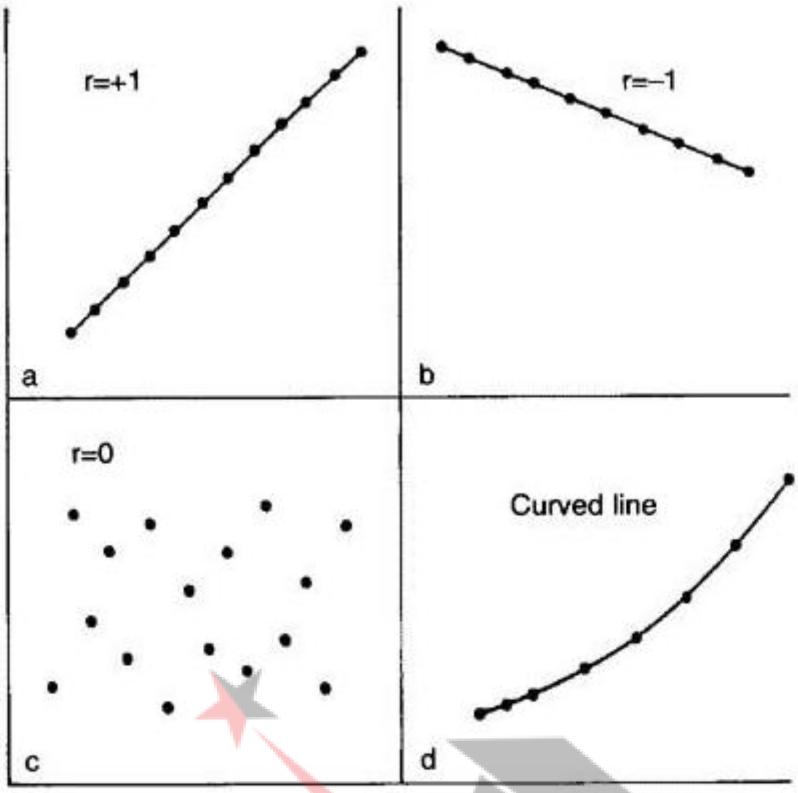
- Correlation shows the quantity of the degree to which two variables are associated. It does not fix a line through the data points. You compute a correlation that shows how much one variable changes when the other remains constant. When r is 0.0, the relationship does not exist. When r is positive, one variable goes high as the other goes up. When r is negative, one variable goes high as the other goes down.
- Linear regression finds the best line that predicts y from x, but Correlation does not fit a line.
- Correlation is used when you measure both variables, while linear regression is mostly applied when x is a variable that is manipulated.

Comparison Between Correlation and Regression

Basis	Correlation	Regression
Meaning	A statistical measure that defines co-relationship or association of two variables.	Describes how an independent variable is associated with the dependent variable.
Dependent and Independent variables	No difference	Both variables are different.
Usage	To describe a linear relationship between two variables.	To fit the best line and estimate one variable based on another variable.
Objective	To find a value expressing the relationship between variables.	To estimate values of a random variable based on the values of a fixed variable.

Correlation and Regression Statistics

The degree of association is measured by “r” after its originator and a measure of linear association. Other complicated measures are used if a curved line is needed to represent the relationship.



The above graph represents the correlation.

The coefficient of correlation is measured on a scale that varies **from +1 to -1 through 0. The complete correlation among two variables is represented by either +1 or -1.** The correlation is positive when one variable increases and so does the other; while it is negative when one decreases as the other increases. The absence of correlation is described by 0.

Correlation Coefficient Formula

If X and Y are two variables, correlation coefficient 'r' is computed as below:

$$r = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y}$$

where $\text{cov}\, X, Y = \frac{1}{N}\sum(x-\bar{x})(y-\bar{y})$

cov (X, Y) is called the covariance between X and Y.

N is the total number of observations.

$\bar{x}$, $\bar{y}$ are the means and σ_x, σ_y are the standard deviations of the variables.

$$\bar{x} = \sum x/N;\ \bar{y} = \sum y/N$$

$$\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$$

$$\sigma_y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{N}}$$

Correlation Coefficient can also be calculated using the formula:

$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\left(\sqrt{N\Sigma x^2 - (\Sigma x)^2}\right)\left(\sqrt{N\Sigma y^2 - (\Sigma y)^2}\right)}$$

Export x	Import y	x^2	y^2	xy
42	56	1764	3136	2352
44	59	1936	3481	2596
58	53	3364	2809	3074
55	58	3025	3364	3190
89	65	7921	4225	5785
98	78	9604	6084	7644
66	58	4356	3364	3828
$\Sigma x = 452$	$\Sigma y = 427$	$\Sigma x^2 = 31940$	$\Sigma y^2 = 26463$	$\Sigma xy = 2846$[illegible]

$$\bar{x} = \frac{452}{7} = 64.57;\ \bar{y} = \frac{427}{7} = 61$$

$$r = \frac{\left(\frac{28469}{7} - 64.57 * 61\right)}{\left(\sqrt{\frac{31940}{7} - 64.57^2}\right) * \left(\sqrt{\frac{26463}{7} - 61^2}\right)}$$

$$r = \frac{(4067 - 3938.77)}{\left(\sqrt{4562.86 - 4169.28}\right) * \left(\sqrt{3780.42 - 3721}\right)}$$

$$r = \frac{(128.23)}{(19.84 * 7.71)}$$

$$r = \frac{128.23}{152.97}$$

Simple Linear Regression Equation

As we know, linear regression is used to model the relationship between two variable. Thus, a simple linear regression equation can be written as:
$Y = a + bX$
Where,

Y = Dependent variable

X = Independent variable

$a = [(\sum y)(\sum x^2) - (\sum x)(\sum xy)] / [n(\sum x^2) - (\sum x)^2]$

$b = [n(\sum xy) - (\sum x)(\sum y)] / [n(\sum x^2) - (\sum x)^2]$