Statistical Modeling

* It's a process of statistical analysis of data set.
* It's usually specified as a mathematical relationship between one or more random variable or non random variable.
* Phases in statistical modeling-
  * Define and Design.
  * Prepare and explore.
  * Refine the model.
  * Answer the research question.

Probability

* Probability its means possibility.
* The meaning of probability is basically the extent to which something is like to happen.

* $\text{Probability} = \frac{\text{Number of way its can happen}}{\text{Total no of outcome}}$

Types of Probabilities

* Classical-there are many possibilities which can equally happen without anything.

e.g. Head of single toss of coin.

* Empirical – its used as theoretical outcome within a limited or controlled field.
* E.g. Tossing of die
* Subjective – Individual person measure the belief that an event will occurred but other person can other.

Distribution

* The distribution of a variable is a description of the relative number of time each possibilities.
* It's a statistical data set is a listing.
* It's a categorical data is organized.
* Percentage of individual in each group by applying various factor.

Types of Distribution

* Normal Distribution-
  * Its also called as symmetric distribution.
  * These types of distribution always happen equally. Or Bell shape curve.
  * The distribution happens by following formula

$$y = \frac{1}{\sqrt{2\pi}}\ e^{-(x-\mu)2/2\sigma}$$

μ =mean σ=standard deviation e=2.71828

Binomial Distribution

* This type of distribution having two possibilities of result i.e success or fail.
* E.g Toss of coin
* Drug Analysis provide two result good or bad.
* Its calculated by
* $P(r)={}_nC_r\,(p)^r\,(1\text{-}p)^{n-r}$
* Mean $\mu = np$
* standard deviation $=\sigma$

Correlation

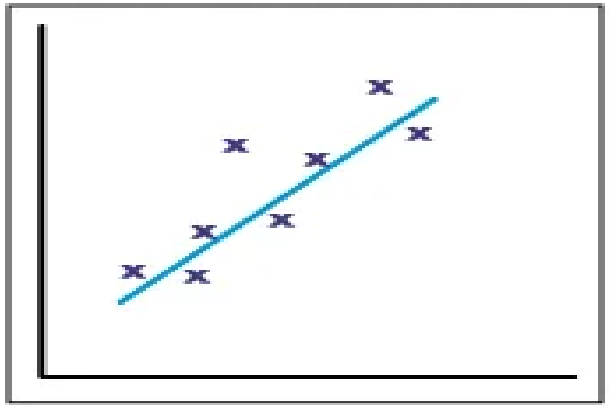
* Correlation is a statistical technique that can show whether and how strongly paired of variables are related.
* Its generally used for measure the linier relationship between in two quantitative variable(or object) .

e.g. Height and weight are related.

Types of Correlation.

* A **positive correlation** is a relationship between two variables in which both variables move in the same direction. Therefore, when one variable increases as the other variable increases, or one variable decreases while the other decreases. An example of positive correlation would be height and weight. Taller people tend to be heavier.
* A **negative correlation** is a relationship between two variables in which an increase in one variable is associated with a decrease in the other. An example of negative correlation would be height above sea level and temperature. As you climb the mountain (increase in height) it gets colder (decrease in temperature).
* A **zero correlation** exists when there is no relationship between two variables. For example there is no relationship between the amount of tea drunk and level of intelligence.

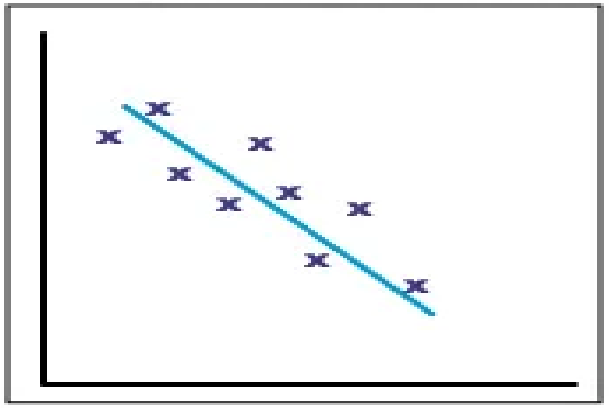
Positive correlation



The points lie close to a straight line, which has a positive gradient.

This shows that as one variable **increases** the other **increases**.

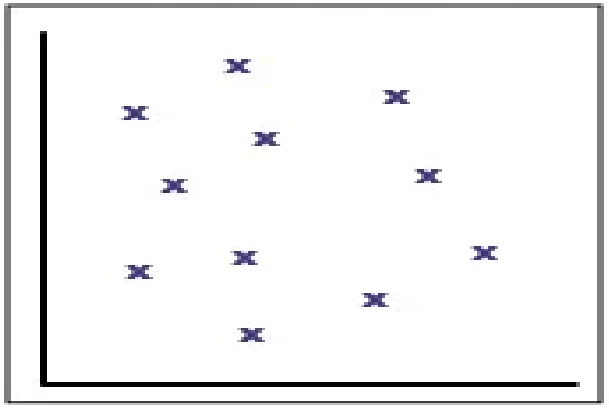
Negative correlation



The points lie close to a straight line, which has a negative gradient.

This shows that as one variable **increases**, the other **decreases**.

No correlation



There is no pattern to the points.

This shows that there is **no connection** between the two variables.