Basic Probability and Statistics

Mentors:
Aditya Pandey
Sanyam Jain

Difference b/w Probability and Statistics

Probability	Statistics
1. Focus: Probability deals with predicting the likelihood of future events based on known information. 2. Application: It is used to study random phenomena and uncertainty, providing a theoretical framework for predicting outcomes. 3. Example: Calculating the probability of rolling a specific number on a fair six-sided die.	1. Focus: Statistics involves collecting, analyzing, interpreting, presenting, and organizing data to make inferences or decisions. 2. Application: It is applied to draw conclusions about populations based on samples, aiding in making informed decisions. 3. Example: Analyzing survey results to make statements about the opinions of a larger population.

Mutually Exclusive and Independent Events

EVENT A AND B ARE MUTUALLY EXCLUSIVE IF:

$$P(A \cap B) = 0$$

EVENT A AND B ARE INDEPENDENT IF:

$$P(A) = P(A \mid B)$$

(APPLYING THE CONDITION B
DOES NOT AFFECT THE PROBABILITY OF A)

Distributions

PDF, PMF and CDF Functions

Random variable:-Variable whose value is unknown to the function.

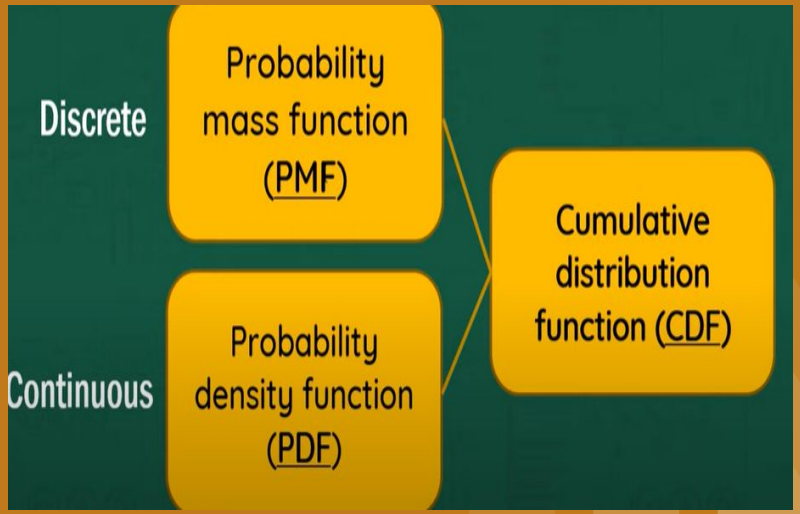
Two types of Random Variable:-

1. Discrete
2. Continuous

PDF(Probability Density Function) is a statistical term that describes the probability distributions of the continuous random variable.

PMF(Probability Mass Function) is a statistical term that describes the probability distributions of discrete random variable.

CDF(Cumulative Distribution Function) is applicable for describing the distribution of random variable either it is discrete or continuous.



Normal Distribution

Definition

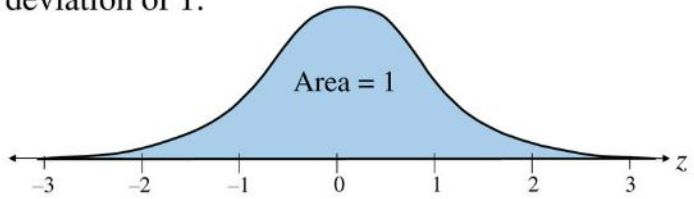
PDF $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Parameter are sigma and Mu.
Central Limit Theorem:- As n(number of outcome) increases, the distribution of sample mean or sum approaches a normal distribution.

Standard Normal Distribution

Standard normal distribution

- A normal distribution with a mean of 0 and a standard deviation of 1.



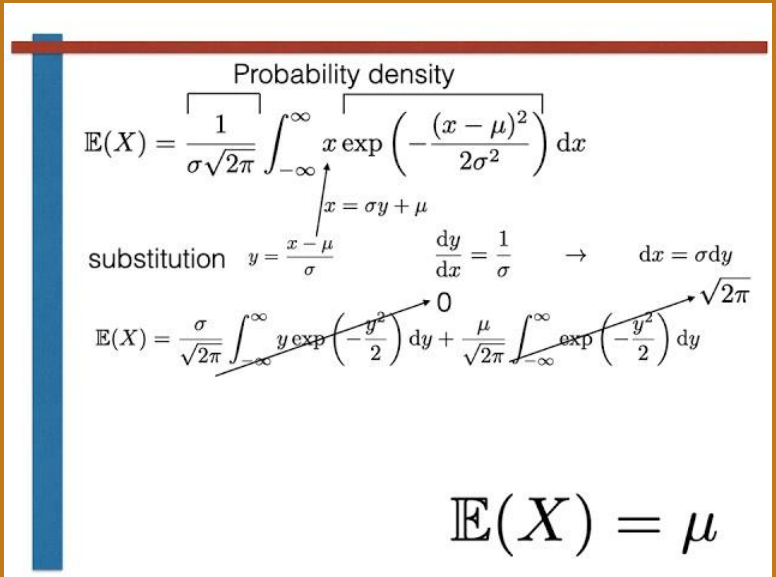
- Any x-value can be transformed into a z-score by using the formula

$$z = \frac{\text{Value} - \text{Mean}}{\text{Standard deviation}} = \frac{x - \mu}{\sigma}$$

Expectation Value

Definition:- The expected value, in statistics and probability, represents the average outcome of a random variable over multiple trials. It is calculated by multiplying each possible outcome by its probability and summing up these values, providing a measure of the central tendency of the variable's distribution.

Expectation value of normal distribution:-



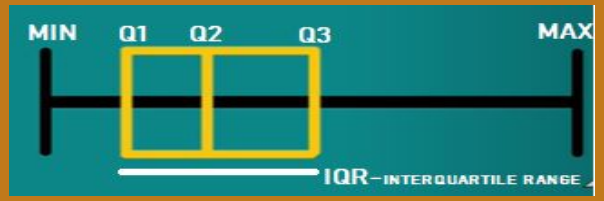
Statistics

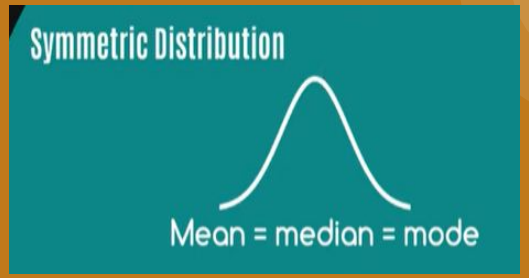
Measures of central tendency: Mean, Median, Mode and their relations under symmetric and asymmetric distributions.

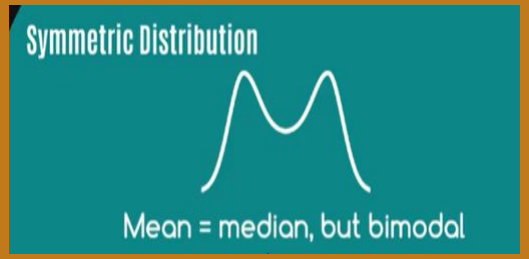
- Mean(Arithmetic): $\Sigma x/n$
- Median: middle number of the ordered series
- Mode: observation with highest frequency
- Mode= 3 x Median - 2 x Mean
- Range= Maximum-Minimum

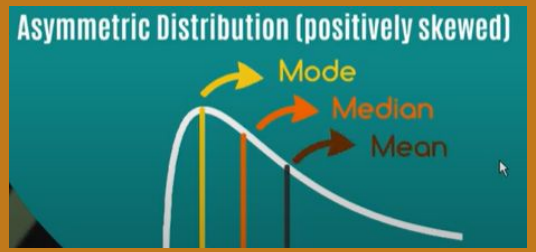
Range and Quartiles

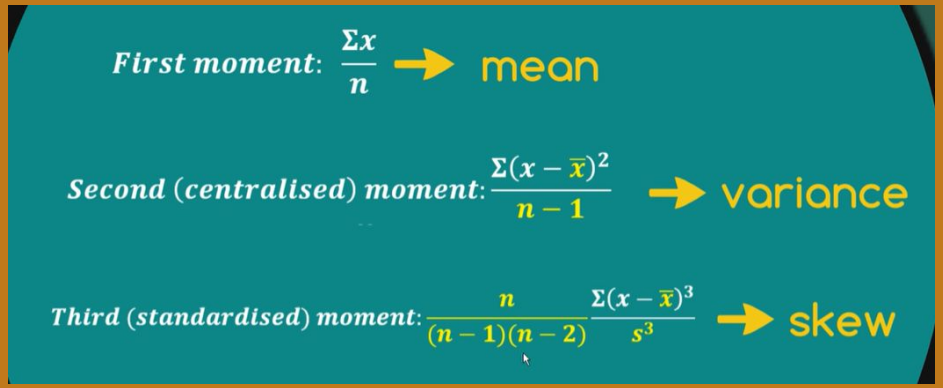
- Quartile- observation 1/4th through the dataset.
  Similarly: Decile 1/10th and Percentile 1/100th











Moments:

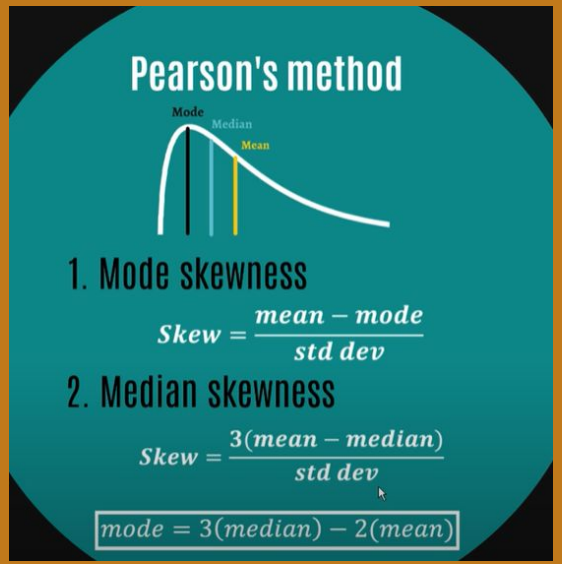
1st => mean

2nd => variance

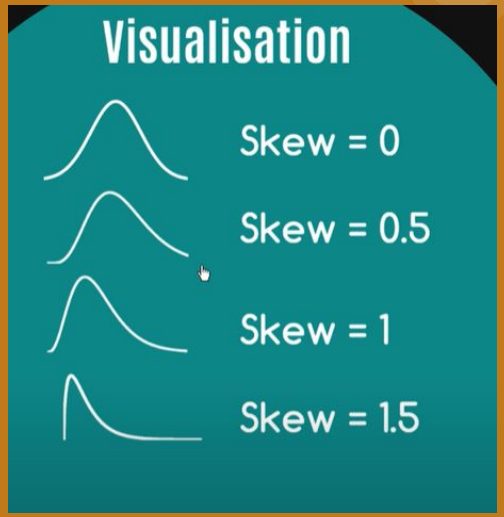
3=> skew

4=> **kurtosis** (peakedness)

Skewness:



- We know that more skewness causes mode, mean and median to separate more.
- Mode and median skewness give a measure of that separation.



Coefficient of variation, covariance and correlation

Coefficient of Variation

$$CV = {}^{s}/_{\overline{x}}$$

Coefficient of variation: takes the scale of the dataset into account.
Standard deviation is in absolute terms, CV is in proportional terms.

Covariance is used to find the direction of the relation between two variables, like if they are directly or inversely proportional

Correlation is used to find the strength of that relation. It is standardized.

$$COV(x,y) = \sigma_{xy} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{n-1}$$

$$CORR(x,y) = \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$