2a)Explain standard deviation & skewness

Standard Deviation :

Standard Deviation is a number used to tell how measurements for a group are spread ou t from the average (mean) or expected value. If we say low standard deviation means mos t of the observations are close to average(mean) normal, if we say high standard deviatio n means most of the observations are spread out (longer distance).

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

standard deviation for sample

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

standard deviation for population

Standard Deviation is the square root of Variance.

Calculating the standard deviation:

Ex : 1,2,3,4,5,6 calculate the standard deviation

Solution:

$$\text{mean} = \frac{1+2+3+4+5+6}{6}$$

mean= 3.5

Here mean is nothing but $\bar{X}$

Now,

Stadard aDeviation = sd = $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$

$$sd = \sqrt{\frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + (x_3-\bar{x})^2 + (x_4-\bar{x})^2 + (x_5-\bar{x})^2 + (x_6-\bar{x})}{6}}$$

$$sd = \sqrt{\frac{(1-3.5)^2+(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2+(6-3.5)^2}{6}}$$

Sd= 1.7078

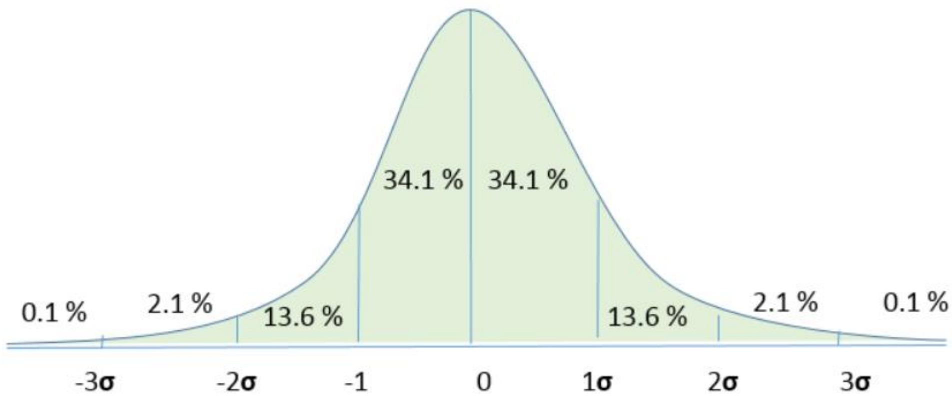
Thumb rule for Standard Deviation

> Below rules are applicable to only Bell shaped data.

1. About 68% of data lie within 1σ of the mean.

2. About 95% of the data lie within 2σ of the mean.

3. About 99.7% of the data lie within 3σ of the mean.

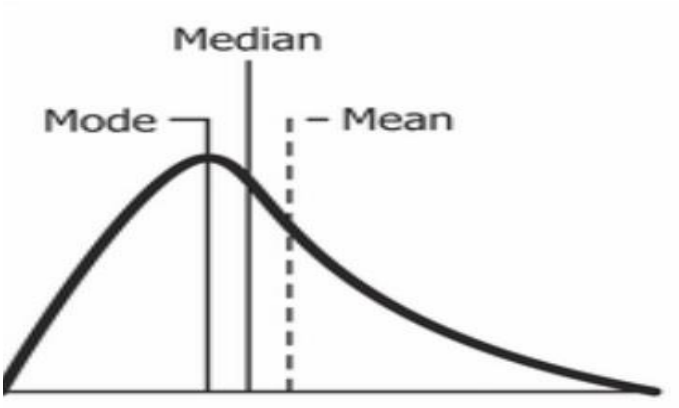


Skewness :

Skewness is a measure of the asymmetry of a distribution. A distribution is asymmetrical when its left and right side are not mirror images.

A distribution can have right (or positive), left (or negative), or zero skewness. A right-skewed distribution is longer on the right side of its peak, and a left-skewed distribution is longer on the left side of its peak

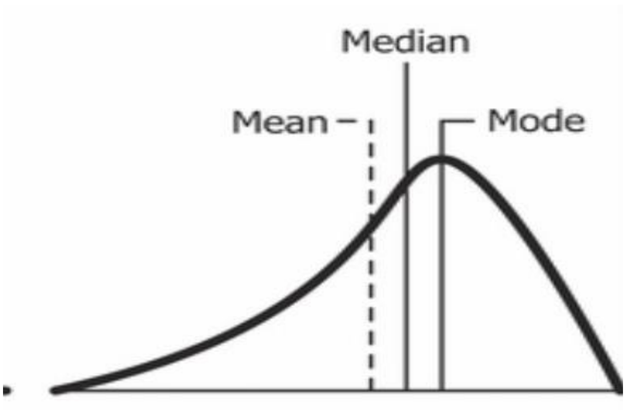
Positive Skewness



For positive skewed curve,

Mode < Median < Mean

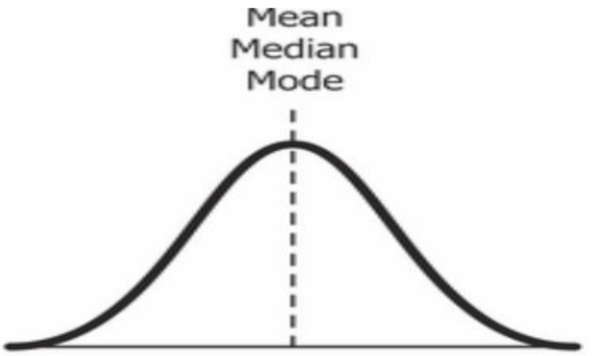
Negative Skewness



For negative skewed curve,

Mean < Median < Mode

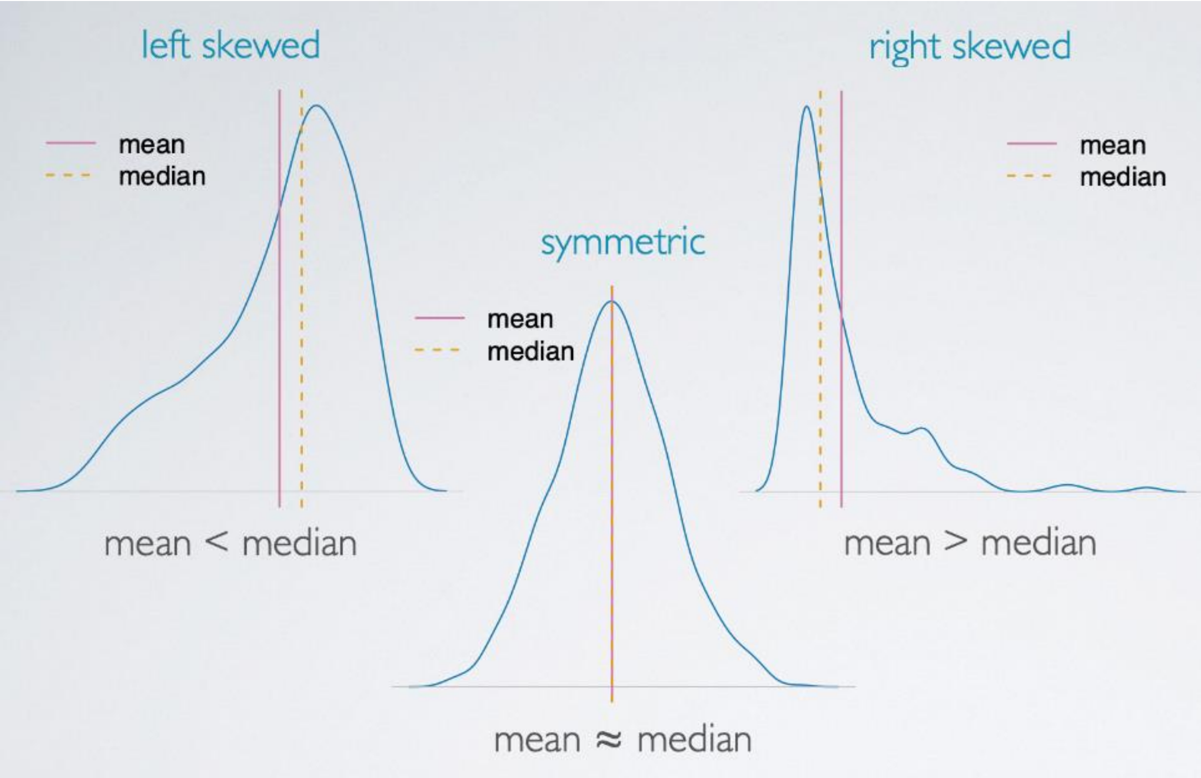
No Skewness



For No Skewed curve,

Mean = Median = Mode

Overview



2b)Explain about Cumulative Probability Distribution &Discrete Probability Distribution.

Cumulative Probability Distribution

In Probability and Statistics, the Cumulative Distribution Function (CDF) of a real-valued random variable, say "X", which is evaluated at x, is the probability that X takes a value less than or equal to the x. A random variable is a variable that defines the possible outcome values of an unexpected phenomenon. It is defined for both discrete and random variables. It is also used to specify the distribution of the multivariate random variables. If the random variable is above a particular level, it is known as tail distribution or the Complementary Cumulative Distribution Function (CCDF). In this article, you will understand what cumulative distribution function is, its properties, formulas, applications and examples.

The Cumulative Distribution Function (CDF), of a real-valued random variable X, evaluated at x, is the probability function that X will take a value less than or equal to x. It is used to describe the probability distribution of random variables in a table.

In other words, CDF finds the cumulative probability for the given value. To determine the probability of a random variable, it is used and also to compare the probability between values under certain conditions. For discrete distribution functions, CDF gives the probability values till what we specify and for continuous distribution functions, it gives the area under the probability density function up to the given value specified.

- The CDF defined for a discrete random variable and is given as

$$F_x(x) = P(X \leq x)$$

- Where X is the probability that takes a value less than or equal to x and that lies in the semi-closed interval (a,b], where a < b.
- Therefore the probability within the interval is written as

$$P(a < X \leq b) = F_x(b) - F_x(a)$$

- The CDF defined for a continuous random variable is given as

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

- Here, X is expressed in terms of integration of its probability density function f_x.
- In case, if the distribution of the random variable X has the discrete component at va

$$P(X = b) = F_X(b) - \lim_{x\to b^-} F_X(x)$$

Examples

The most important application of cumulative distribution function is used in statistical analysis. In statistical analysis, the concept of CDF is used in two ways.

- Finding the frequency of occurrence of values for the given phenomena using cumulative frequency analysis.
- To derive some simple statistics properties, by using an empirical distribution function, that uses a formal direct estimate of CDFs.

Note:

The cumulative distribution function (CDF) calculates the cumulative probability for a given x-value. Use the CDF to determine the likelihood that a random observation taken from the population will be less than or equal to a particular value.

Discrete Probability Distribution

A discrete probability distribution counts occurrences that have countable or finite outcomes.This is in contrast to a continuous distribution, where outcomes can fall anywhere on continuum.Common examples of discrete distribution include the binomial, Poisson, and Bernoulli distributions. These distributions often involve statistical analyses of "counts" or "how many times" an event occurs. In finance, discrete distributions are used in options pricing and forecasting market shocks or recessions.

Distribution is a statistical concept used in data research. Those seeking to identify the outcomes and probabilities of a particular study will chart measurable data points from a data set, resulting in a probability distribution diagram. There are many types of probability distribution diagram shapes that can result from a distribution study, such as the normal distribution ("bell curve").

There are many discrete probability distributions to be used in different scenarios. We will discuss Discrete distributions in this post. Binomial and Poisson distributions are the most discussed ones in the following list.

1. ***Bernoulli Distribution***

$$\text{PMF} = \begin{cases} \text{p}, & \text{Success} \\ 1 \text{ - p}, & \text{Failure} \end{cases}$$

2. ***Binomial Distribution***

$$n_{C_x} p^x (1-p)^x$$

3. Hypergeometric Distribution

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

4. Negative Binomial Distribution

$$P(X = k) = \binom{k + r - 1}{k} p^{r}(1-p)^{k}$$

5. Geometric Distribution

$$P(X = k) = (1-p)^{k-1}\, p$$

6. Poisson Distribution

$$P(X = x) = \frac{e^{-\lambda}\, \lambda^{x}}{x!}$$

7. Multinomial Distribution

$$P(X = x_1, X = x_2, \ldots. X = x_k) = \frac{n!}{x_1!\, x_2! \ldots . x_k!}\, {p_1}^{x_1}\, {p_2}^{x_2} \ldots . {p_k}^{x_k}$$