__/__/__

Measure of Central Tendency

Important Formulae :

Arithmetic Mean:

Individual Series:

1. $\overline{X} = \frac{\Sigma X}{N}$ (Direct Method)

2. $\overline{X} = A + \frac{\Sigma d}{N}$ (Short-cut Method); where, $d = (X - A)$

3. $\overline{X} = A + \frac{\Sigma d'}{N} \times i$ (step-deviation Method); $d = \left(\frac{X - A}{i}\right)$

Discrete Series:

1. $\overline{X} = \frac{\Sigma fx}{N}$ (Direct Method)

2. $\overline{X} = A + \frac{\Sigma fd}{N}$ (Short-cut Method); $[d = (X - A)]$

3. $\overline{X} = A + \frac{\Sigma fd'}{N} \times i$ (step-deviation Method); $\left[d = \frac{X - A}{i}\right]$

Continuous Series:

1. $\overline{X} = \frac{\Sigma fm}{N}$ (Direct Method); m = (mid value)

2. $\overline{X} = A + \frac{\Sigma fd}{N}$ (Short-cut Method); $[d = m - A]$

3. $\overline{X} = A + \frac{\Sigma fd'}{N} \times i$ (Step-deviation method); $\left[d' = \left(\frac{m-A}{i}\right)\right]$

♦ **Weighted Arithmetic Mean:**

$$\overline{X}_W = \frac{\Sigma WX}{\Sigma W}$$

♦ **Combined Mean:**

$$\overline{X}_{123} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2 + N_3\overline{X}_3}{N_1 + N_2 + N_3}$$

♦ **Median:**

Individual Series:

M = Size of $\left(\frac{N+1}{2}\right)$th item

Discrete Series:

M = Size of $\left(\frac{N+1}{2}\right)$th item; $[N = \Sigma f]$

Continuous Series:

$M = l_1 + \frac{\frac{N}{2} - c.f}{f} \times i$; M = Size of $\left(\frac{N}{2}\right)$th item

l_1 = lower limit of the median class
f = frequency of median class
$C \cdot f$ = Commulative frequency
i = size of median class

Mode:

Individual Series / Discrete Series:

$$Z = (\text{maximum frequent number in the series})$$

Continuous Series:

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

f_1 = frequency of modal class
f_0 = frequency of pre-modal class
f_2 = frequency of post-modal class

*** If the modal value lies outside the modal class:

$$Z = l_1 + \frac{f_2}{f_0 + f_2} \times i$$

__/__/____

♦ Emprical Formula:

$$Mode = 3Median - 2Mean$$

'or'

$$Z = 3M - 2\bar{X}$$

♦ Geometric Mean:

Individual Series:

$$GM = Antilog\left[\frac{\Sigma \log x}{N}\right]$$

Discrete Series and Continuous Series:

$$GM = Antilog = \left[\frac{\Sigma f \log x}{N}\right]$$

Measure Of Despersion

Range:

$$Range = L - S$$

$$\text{Coefficient Of Range} = \frac{L-S}{L+S}$$

Quartile Deviation:

- Individual And Descrete Series:

$Q_1 = \left(\frac{N+1}{4}\right)th$ • Quartile Deviation $= \frac{Q_3 - Q_1}{2}$

$Q_3 = \frac{3(N+1)th}{4}$ • Coefficient of Q.D $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

- Inter Quartile Range $= Q_3 - Q_1$

- Continuous Series:

- $Q_1 = \left(\frac{N}{4}\right)$th item • Quartile Deviation $= \frac{Q_3 - Q_1}{2}$

- $Q_3 = \frac{3N}{4}$ th item • Coefficient of Q.D $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

- Inter Quartile Range $= Q_3 - Q_1$

Mean Deviation:

- Individual Series:

$M.D = \frac{\Sigma|D|}{N}$ or $\frac{\Sigma|X - \bar{X}|}{N}$ [Mean Deviation from Mean]

$M.D = \frac{\Sigma|X - M|}{N}$ [Mean Deviation from Median]

_/_/_

- **Discrete and Continuous Series:**

$$M \cdot D = \frac{\Sigma f |D|}{N} \text{ or } \frac{\Sigma f |x - \bar{x}|}{N}; \text{[M·D from mean]}$$

$$M \cdot D = \frac{\Sigma f |X - M|}{N}; \text{[M·D from median]}$$

$$\text{Coefficient of } M \cdot D = \frac{M \cdot D}{\text{Average}} \text{ or } \frac{M \cdot D}{\text{Median}}$$

Standard Deviation:

- **Individual series:**

1. $\sigma = \sqrt{\frac{\Sigma (X - \bar{X})^2}{N}}$ (Actual Mean method)

2. $\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$ (Assumed Mean method); $d = X - A$

3. $\sigma = \sqrt{\frac{\Sigma x^2}{N} - \left(\frac{\Sigma x}{N}\right)^2}$ [Method based on use of actual data]

- **Discrete Series:**

1. $\sigma = \sqrt{\frac{\Sigma f (x - \bar{x})^2}{N}}$ (Actual Mean Method)

2. $\sigma = \sqrt{\frac{\Sigma f d^2}{N} - \left(\frac{\Sigma f d}{N}\right)^2}$; $d = X - A$ (Assumed mean method)

3. $\sigma = \sqrt{\frac{\Sigma f d'^2}{N} - \left(\frac{\Sigma f d'}{N}\right)^2} \times i$; $d' = \frac{X - A}{i}$;

(step Deviation method)

__/__/__

- **Continuous Series:**

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i$$

$$d' = \left(\frac{m-A}{i}\right)$$

- Coefficient of Standard Deviation $= \dfrac{\sigma}{\overline{X}}$

Variance :

- Variance $= \dfrac{\Sigma f(x-\bar{x})^2}{N}$ (Actual Mean Method)
- Variance $= \left[\dfrac{\Sigma fd^2}{N} - \left(\dfrac{\Sigma fd}{N}\right)^2\right]$ (Assumed Mean method)
- Variance : $\left[\dfrac{\Sigma fd'^2}{N} - \left(\dfrac{\Sigma fd'}{N}\right)^2\right] \times i$ (Step-Deviation method)
- Coefficient of Variation $= \dfrac{\sigma}{\overline{X}} \times 100$
- Variance $= (S.D)^2$

—/—/—

Combined Standard Deviation:

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1+N_2}}$$

$$\left(d_1 = \overline{X}_1 - \overline{X}_{12}\right), \left(d_2 = \overline{X}_2 - \overline{X}_{12}\right)$$

$$\left(\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1+N_2}\right)$$

__/__/__

Sampling

Sampling: The actions or process of taking samples of somthing from large population for analysis.

Difference between population and sample.

A population is the entire group that you want to draw conclusions about. A sample is the specific group that you will collect data from.
The size of sample is always less than the size of population.

Random sample:

Random sampling is a part of the sampling technique in which each sample has an equal probability of being choosen. A sample chosen randomly is meant to be unbiased representation of the total population.

Sampling Distribution of a statistics:

A sampling distribution refers to a probability distribution of a statistics that comes from choosing random samples of a given population.

Sampling Distribution of mean :

The sampling distribution of mean is normally distributed. This means, the distribution of sample means of large sample size is normally distributed irrespective of the shape of the universe, but provided the population standard deviation (σ) is finite.

1. The mean of sampling distribution of mean is equal to the population mean (μ)

$$\bar{X} = \mu$$

2. The standard error of the sampling distribution of means is obtained as:

$$SE_{\bar{X}}(\sigma_{\bar{X}}) = \frac{\text{S.D of population}}{\sqrt{\text{Size of sample}}} = \frac{\sigma}{\sqrt{n}}$$

$$Var(\bar{X}) = \frac{\Sigma f(\bar{X} - \mu_{\bar{x}})^2}{\Sigma f}$$

$$Var(\bar{x}) = \frac{\sigma^2}{n}$$

__/__/__

3. probability of the sampling distribution of means.

$$Z = \frac{\overline{x} - \mu}{\sigma}$$

** With replacement:
Total no. of sample $= N^n$; (N: Population size)

** Without replacement:
Total no. of sample $= {}^{N}C_{n}$ (n = Sample size)

Test of Hypothesis:

In attempting to arrive at decisions about the population on the basis of sample information it is necessary to make assumptions about the population parameters involved. Such an assumption is called a statistical hypothesis, which may or may not be true.

→ Null Hypothesis:

In test of hypothesis, we always begin with an assumption or hypothesis. This is called Null Hypothesis.

There is no (significant) difference b/w statistics and population parameter,

$$H_0 : (\mu = \overline{x})$$

—/—/—

→ Alternative Hypothesis :

Any hypothesis different from NULL hypothesis (H_o) is called an alternative hypothesis (H_1).

H_1 : $\mu \neq \mu_o$

$\mu > \mu_o$

$\mu < \mu_o$

→ Type 1 Error :

It is made when we reject the NULL hypothesis (H_o), although it is true

Ex: Drug.

→ Type 2 Error:

It is made when we accept Null hypothesis which we accept and is although false.

→ Level of significance:

This refers to degree of significance with which we accept and reject the hypothesis.

→ Critical Region or Rejection Region :
The critical region or rejection region is the region of the standard normal curve corresponding to pre-determined level of significance. The region under the normal curve which is not covered by the rejection region is known as acceptance region. Thus, the statistic which leads to the rejection of null hypothesis H_0 gives us a region, known as critical region.

→ 1-Tailed Test and 2-Tailed Test :
A test of any statistical hypothesis where the alternative hypothesis is expressed by symbol ($<$) or the symbol ($>$) is called a one-tailed test since the entire critical region lies in one tail of the distribution of the test statistics.

A test of any statistical hypothesis where the alternative is written with symbol "$\neq$" is called a two-tailed test.

_/_/_

Procedure for Testing Hypothesis:

Step 1: Set up Null Hypothesis
Step 2: Alternate hypothesis
Step 3: Level of significance
Step 4: Calculation
Step 5: Conclusion

Test of Hypothesis for large Sample (n>30):

1. Test of hypothesis about population mean:

$$Z = \frac{\bar{X}-\mu}{S.E_{\bar{x}}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

$|Z| < Z_\alpha$: NH (H_0) accepted and AH (H_1) Rejected

$|Z| > Z_\alpha$: NH (H_0) Rejected and Accepted AH (H_1)

1% LOS

One Tailed $Z_\alpha = 2.33$
Two " $Z_\alpha = 2.58$

5% LOS

1. $Z_\alpha = 1.96$ (Two tailed)
2. $Z_\alpha = 1.64$ (One-tailed)

__/__/__

2. Test of hypothesis about difference between population mean:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{SE_{\bar{X}_1 - \bar{X}_2}}$$

$$SE_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

or

$$SE_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$H_0 : \mu_1 = \mu_2$ (There is no diff. b/w two population mean)

$H_1 : \mu_1 \neq \mu_2$ → {Two-tailed}

or $\underbrace{H_1 : \mu_1 > \mu_2 \text{ or } \mu_1 < \mu_2}_{\text{(One tailed)}}$

__/__/__

3. Test of Hypothesis ~~for~~ about difference between two population standard deviation:

$$Z = \frac{(S_1 - S_2) - (\sigma_1 - \sigma_2)}{SE_{S_1 - S_2}}$$

$$SE_{S_1 - S_2} = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}} \quad \text{or} \quad \sqrt{\frac{S_1^2}{2n_1} + \frac{S_2^2}{2n_2}}$$

$H_0 : \sigma_1 = \sigma_2$ (there is no diff. b/w two population SDs)

$H_1 : \sigma_1 \neq \sigma_2$

or $\sigma_1 > \sigma_2$ or $\sigma_1 < \sigma_2$

_/_/_

Test of Hypothesis for small Sample Test:

1. t-test:

$$t = \frac{\text{Deviation from the population parameter}}{\text{Standard Error of sample statistic}}$$

- **Test of hypothesis about the population mean:**
  (σ unknown, sample size small)

  - $$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

    S = modified S.D of the sample.

  - $$S = \sqrt{\frac{\sum (X - \bar{X})^2}{(n-1)}} \quad \text{(Actual mean)}$$

    $$S = \sqrt{\frac{\sum d^2 - (\bar{d})^2 \times n}{n-1}} \quad , \quad d = X - A \ \text{(Assumed mean)}$$

    $$S = \sqrt{\frac{n}{n-1} \cdot s^2} \quad \text{(standard deviation}$$

- Degree of freedom $= \nu = n - 1$

__/__/__

- Test of hypothesis about difference between two means in case of independent sample:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$S = \sqrt{\frac{\Sigma(x_1 - \bar{x}_1)^2 + \Sigma(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}}$$

$$S = \sqrt{\frac{\Sigma(X_1 - A_1)^2 + \Sigma(X_2 - A_2)^2 - n_1(\bar{X}_1 - A_1)^2 - n_2(\bar{X}_2 - A_2)^2}{n_1 + n_2 - 2}}$$

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

- Degree of freedom $= \nu = n_1 + n_2 - 2$

- Test of hypothesis about diff. b/w two means in case of dependent:

$$t = \frac{\bar{d}}{S} \cdot \sqrt{n}$$

$$S = \sqrt{\frac{\Sigma d^2 - (\bar{d})^2 \times n}{n-1}}$$

$d.f = \nu = n - 1$

__/__/__

2. F-Test (Variance Ratio Test)

$$F = \frac{\text{Larger estimate of population variance}}{\text{Smaller estimate of population Variance}} = \frac{S_1^2}{S_2^2}$$

where ; $S_1^2 > S_2^2$

Procedure :

1. Set up null hypothesis i.e. $H_0 : \sigma_1^2 = \sigma_2^2$.

2. Calculation :

$$S_1^2 = \frac{\sum(X_1 - \bar{X}_1)^2}{n_1 - 1} \quad \text{or} \quad \frac{n_1}{n_1 - 1} \cdot s_1^2 \quad \text{or} \quad \frac{1}{n_1 - 1}\left[\sum x_1^2 - \frac{(\sum(x_1))^2}{n_1}\right]$$

Sample Variance.

$$S_2^2 = \frac{\sum(X_2 - \bar{X}_2)^2}{n_2 - 1} \quad \text{or} \quad \frac{n_2}{n_2 - 1} \cdot s_2^2 \quad \text{or} \quad \frac{1}{n_2 - 1}\left[\sum x_2^2 - \frac{(\sum x_2)^2}{n_2}\right]$$

Sample variance,

3. Ratio :

$$F = \frac{S_1^2}{S_2^2} \; ; \; S_1^2 > S_2^2$$

4. $\nu_1 = n_1 - 1$
   $\nu_2 = n_2 - 1$

_/_/_

5. LOS (5%) in F-table.

6. $|F| < F_\alpha \rightarrow$ Accept NH, Reject AH
$|F| > F_\alpha \rightarrow$ Reject NH, Accept AH

Chi(χ^2)-Square Test:

1. χ^2 - test as a test for population variance.
2. χ^2 - test as a non-parametric test.

1. # χ^2-test as a test for population Variance:

Procedure:

1. Set up the null hypothesis:

$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_1 : \sigma^2 > \sigma_0^2$$

2. Compute χ^2 :

$$\chi^2 = \frac{\Sigma(x-\bar{x})^2}{\sigma^2} \text{ or } \frac{ns^2}{\sigma^2}$$

where, $s^2 = \frac{\Sigma(x-\bar{x})^2}{n}$

$$ns^2 = \Sigma(x-\bar{x})^2$$

3. Degree of freedom : $\nu = n-1$

4. LOS

5. Conclusion .

$\chi^2 > \chi^2_\alpha$; Reject NH & Accept AH

$\chi^2 < \chi^2_\alpha$; Accept NH & Reject AH.

2. # χ^2-test as a non-parametric test :

(i) χ^2- test as a test of goodness of fit:

Procedure :

1. Set up null hypothesis : H_0 :
2. Compute χ^2 :

$$\chi^2 = \Sigma \left[\frac{(O-E)^2}{E} \right]$$

O = Observed Frequency
E = Expected Frequency

$$\chi^2 = \Sigma \left[\frac{O^2}{E} \right] - N$$

N: total expected frequency , $\Sigma O = \Sigma E = N$

__/__/__

3. degree of freedom: $\nu = n-1$

Binomial: $\nu = n-1$
Poisson : $\nu = n-2$
Normal : $\nu = n-3$

4. The calculated value of χ^2 as such is than Compared with the table value of χ^2 for give degree of freedom at 5% and 1% LOS. If the calculated χ^2 exceeds the table value of χ^2. We reject H_0 and conclude that the fit is not good. if χ^2 less than table χ^2, then ~~the~~ we accept H_0 and conclude the fit is good.

Expected frequency $= np$

n = Sum of observations
p = probability of each observations.

2. χ^2 - test as a test of independence of attributes:

Procedure:

1: Set null hypothesis: two attributes are independent which means two means are associated or independent of one another.

2. Calculate Expected Frequency,

$$\text{Expected frequency} = \frac{R \times C}{N}$$

R = Row Total

C = Column Total

N = Total no. of Observation

3. Compute χ^2:

$$\chi^2 = \Sigma \left[\frac{(O-E)^2}{E} \right]$$

4. Degree of freedom: $\nu = (c-1)(r-1)$

5. LOS

6. Conclusion.

__/__/__

Q. What is difference between parametric and non-parametric Test ?

Ans:-

Parametric	Non-Parametric
→ Makes assumptions about the population	→ Doesn't make any assumptions.
→ Use a normal probabilistic distribution.	→ The distribution is arbitrary.
→ Uses mean value for central tendency.	→ Uses median value for central tendency.
→ Requires previous knowledge about the population.	→ Doesn't require previous knowledge about the population.

- Non-Parametric χ^2 test

It is used for testing the significance of a population variance. It is a non-parametric test, and it can be used as test of goodness of fit and Independent of attribute.

_/_/_

Regression

It is the study of nature of relationship between the variables; so that one may be able to predict the unknown value of one variable for a known value of another variable to stop.
It is average relationship between two variables.

Utilities of Regression:

1. Nature of relationship
2. Estimation of relationship
3. Prediction of relationship
4. It is useful in economics and business research
5. Correlation is a measure of degree of relationship b/w x and y variable wherease regression is a nature of relationship b/w two variables.

Equation of Regression line: (Avg. relationship b/w Variables)

$$Y = a + xb \quad \left[\text{Equation of } Y \text{ on } X\right]$$

or

$$X = a + Yb \quad \left[\text{Equation of } X \text{ on } Y\right]$$

_/_/_

Regression Equation:

It is algebraic formulation of regression line.

Y on X (Regression Equation Y on x):

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$$Y = \bar{Y} + b_{yx}(X - \bar{X})$$

$$\boxed{b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}}$$

Regression coefficient

Regression Coefficient:

Measure the slope of regression line.
It measure the average change in the value of one variable for a unit change in another variable.

Properties:

1. $b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y}$ and $b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}$

$$\boxed{b_{xy} \cdot b_{yx} = r}$$

2. $r = \pm 1$

Calculation of Regression Equation:

→ Individual Series:

Two method:

1. Normal Equation (Least Square Equation):

$$Y = a + bx$$

$$\boxed{\Sigma y = Na + b\Sigma x}$$

$$\boxed{\Sigma xy = a\Sigma x + b\Sigma x^2}$$

$$X = a + bY$$

$$\Sigma x = Na + b\Sigma y$$

$$\Sigma xy = a\Sigma y + b\Sigma y^2$$

2. Regression Equation using regression coefficient:

(a) Using Actual Value of X and Y:

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$$Y = \bar{Y} + b_{yx}(X - \bar{X})$$

$$b_{yx} = \frac{N\Sigma xy - \Sigma x \cdot \Sigma y}{N\Sigma x^2 - (\Sigma x)^2}$$

Probability

- **Experiment:** When we conduct a trial to obtain statistical information, it is called an experiment.

Ex:- Tossing of a fair coin is an experiment, it has two possible outcomes.

- **Events:** The possible outcomes of a trial/experiment are called Events.

Ex:- If a fair coin is tossed, the outcomes: head or tail are called events.

- **Probability:** It is the ratio of the favourable cases to the total number of equally likely cases.

$$P(A) = p = \frac{\text{No. of favourable cases}}{\text{Total no. of Equally likely cases}} = \frac{m}{n}$$

Similarly,

$$P(\overline{A}) = q = 1 - p(A) = 1 - \frac{m}{n}$$

__/__/__

- **Equally likely Event :**
If the events have the same theoritical ~~formula~~ probability of happening then they are called Equally likely Events.

- **Complementary Event :**
In the case of such events, there will only be two outcomes that states whether an event will occur or not.

- **Mutually Exclusive events :**
Mutually Exclusive Events are those events that do not occur at the same time.

- **Exhaustive Event :**
Exhaustive Events are the events obtained when a sample space is broken down into events that are mutually exclusive in nature such that their union forms the sample space itself.

_/_/_

- Use of combinations in Theory probability:

$$^nC_r = \frac{\lfloor n}{\lfloor n-r \cdot \lfloor r}$$

Theorems of Propability:

1.) Addition Theorem
2.) Multiplication Theorem
3.) Baye's Theorem.

1. Addition Theorem:

> Addition Theorem for Mutually Exclusive Events:

$$P(A \text{ or } B) = P(A) + P(B)$$
$$P(A+B) = P(A) + P(B)$$

Generalization:

$$P(A+B+C) = P(A) + P(B) + P(C)$$

> Addition Theorem for not Mutually Exclusive Events:

$$P(A \text{ or } B \text{ or both}) = P(A) + P(B) - P(AB)$$

$$P(\text{Either } A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) - P(ABC)$$

__/__/__

2. Multiplication Theorem :

> Multiplication Theorem for Independent Events :

$$P(AB) = P(A) \times P(B)$$

Generalization :

$$P(ABC) = P(A) \times P(B) \times P(C)$$

Conditional Probability :

The probability of occurance of event A given that event B has already occured is known as conditional probability.

$$P(A/B) = \frac{n(A \cap B)}{n(B)}$$

Or $$P(B/A) = \frac{n(A \cap B)}{P(A)}$$

- $P(A \cup B/C) = P(A/C) \cup P(B/C) - P(A \cap B/C)$
- $P(A'/B) = 1 - P(A/B)$

> Multiplication Theorem for Dependent Events :
or in case of Conditional probability :

$$P(AB) = P(A) \cdot P(B/A)$$

or $$P(AB) = P(B) \cdot P(A/B)$$

_/_/_

Bernoulli's Theorem in Theory of probability:

$$P(r) = {}^{n}C_{r} \cdot p^{r} \cdot q^{n-r} \quad , \quad r : 1, 2, 3, \ldots, n$$

$P(r)$ = Probability of r successes in n trials
p = probability of success or happening of an event in one trial
q = probability of failure.
n = Total no. of trials

Baye's Theorem:

It was given by British mathematecian Thomas Bayes in 1763.

$$P(A_1) \longrightarrow P(B/A_1) \longrightarrow P(A_1) \times P(B/A_1)$$

$$P(B_2) \longrightarrow P(B/A_2) \longrightarrow P(A_2) \times P(B/A_2)$$

$$P(A_1/B) = \frac{P(A_1) \times P(B/A_1)}{P(A_1) \times P(B/A_1) + P(A_2) \times P(B/A_2)}$$

Probability Distribution

1. Observed Frequency Distribution
2. Theoritical Frequency Distribution

1. Observed Frequency Distribution :
Observed Frequency Distribution refers to those frequency distributions which are obtained by actual observation or experiments.

2. Theoritical or Probability Distribution :
Theoritical frequency distribution refers to those distributions which are not obtained by actual observations or experiments but are mathematically deduced under certain assumptions.

Types of Probability Distribution

1. Binomial Distribution
2. Poisson Distribution
3. Normal Distribution

_/_/_

1. Binomial Distribution :

- Mean$(\bar{x}) = np$
- Variance $(\sigma^2) = npq$
- S.D $(\sigma) = \sqrt{npq}$
- Moment coeff. of skewness $= \sqrt{\beta_1} = \dfrac{q-p}{\sqrt{npq}}$
- Moment coeff. of Kurtosis $= (\beta_2) = 3 + \dfrac{1-6pq}{npq}$
- $P(r) = {}^nC_r \cdot p^r \cdot q^{n-r}$

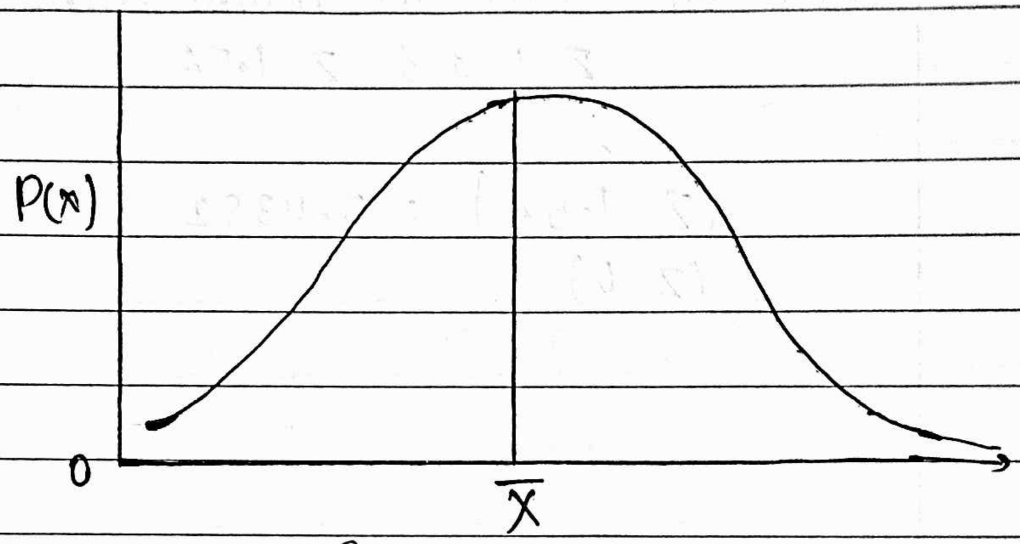
2. Poisson Distribution :

- $p(X=x) = \dfrac{e^{-\lambda} \lambda^x}{x!}$
- Mean $= \lambda = \bar{X} = np$
- Variance $= \sigma^2 = \lambda$
- S.D $= \sigma = \sqrt{\lambda}$
- Moment of skewness $= \sqrt{\beta_1} = \dfrac{1}{\sqrt{\lambda}}$
- Moment of Kurtosis $= \beta_2 = 3 + \dfrac{1}{m}$

- **Normal Distribution:**

  - $p(x) = \frac{1}{\sigma\sqrt{2\pi}}\left[e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{2}\right)^2}\right]$; $-\infty < X < \infty$

$\bar{X}$: Mean , σ : S.D



Graph of Normal Distribution

- ($\bar{X} = M = Z$)
- Unimodal Distribution
- Perfectly Symmetrical and Bell shaped
- Asymptotic to the Bell Line
- Range : $-\infty$ to ∞
- Total Area under the curve is 1
- $Q.D = \frac{2}{3} S.D$
- $M.D = \frac{4}{5} S.D$

Area Under the normal Curve:

$$Z = \frac{X - \bar{X}}{\sigma}$$

Example:

Find Area Under the normal curve between

$Z = 0$ and $Z = 1.54$

$(Z = 1.54) = 0.4382$

$(Z = 0)$

__/__/__

Spearman Rank Correlation:

$$R = 1 - \frac{6\Sigma D^2}{N^3 - N}$$

D = Difference b/w two ranks

N = total no. of ranks

When equal or Tied Rank:

$$R = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)\right]}{N^3 - N}$$

m = No. of repeatations