① What is statistics?

A mathematical discipline to collect and summarize data. Collection, analysis, interpretation, presentation & organizing a data is called statistics.

② What is probability?

This is the measure of the likelihood of an event to happen.

$$P(E) = \frac{\text{no. of favourable outcomes}}{\text{no. of total outcomes}}$$

③ What are the types of statistics?

There are two types.

- → ~~Measure~~ Descriptive statistics
- → Inferential statistics.

Descriptive Statistics describes the properties of sample and population.
Inferential statistics use those properties to test hypotheses and draw conclusion

Descriptive statistics is divided into two types

- → Measure of central tendency
- → Measure of variability

Date

* Measure of central tendency is a summary measure, that attempts to describe the whole data with a single value, that represents the middle or centre of distribution

- → mean
- → median
- → mode

* Measure of variability describes how the data is spread around the mean.

- → Range
- → Variance
- → Standard deviation.

(4) **Mean:** Average of the data.

i.e., $\bar{x} = \dfrac{\text{Sum of observations}}{\text{Total no. of observations.}}$

Median: When arranged in ascending or descending order, the middle value is called median.

i.e., $\mu = \left(\dfrac{n+1}{2}\right)$th term for $n =$ odd

$$\left\{\frac{\left(\frac{n}{2}\right) + \left(\frac{n+1}{2}\right)}{2}\right\} \text{ for } n = \text{even}$$

mode: Frequently occurred value in the data.

Examples:—

~~data set = {10, 12, 13, 14, 18, 17, 20, 25, 40, 48, 50, 56, 20, 28, 14,~~

data set = {10, 12, 12, 16, 18, 20, 20, 20, 25, 40, 40, 48}

$$\bar{x} = \frac{\text{Sum of observations}}{\text{Total no. of observations}}$$

$$\Rightarrow \frac{337}{13}$$

mean $(\bar{x}) = 25.923$

median $(\mu) = \frac{n+1}{2}$ th term) $\frac{13+1}{2} \Rightarrow \frac{14}{2}$ th term $\Rightarrow 7$ th term

$\therefore$ median $(\mu) = 20$

mode $= 20$.

$\therefore$ mean $(\bar{x}) = 25.923$
median $(\mu) = 20$
mode $= 20$

Page No.
Date

(5) Qualitative and Quantitative analysis?

Analysis done for ~~Qua~~ Categorical variables is called Qualitative analysis. That means, analysis is done on the data, based upon its properties, or behaviour

Eg: Gender, Race, etc.

Analysis done on the numerical data is called Quantitative analysis.

Eg Height, weight, age etc.

(6) Frequency distribution

This represents, how frequently the particular data is repeated in the given range. It can be in the form of a graph or table.

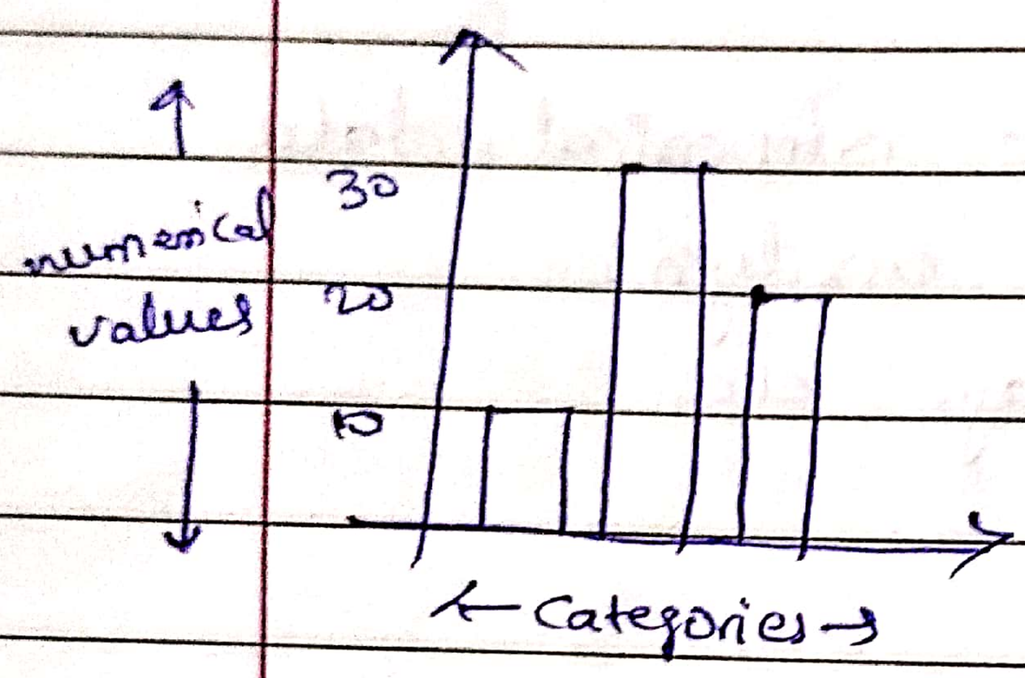
For example, consider scores of a students from a class with 9 students.

Scores = {65, 70, 71, 72, 75, 78, 80, 85, 95}

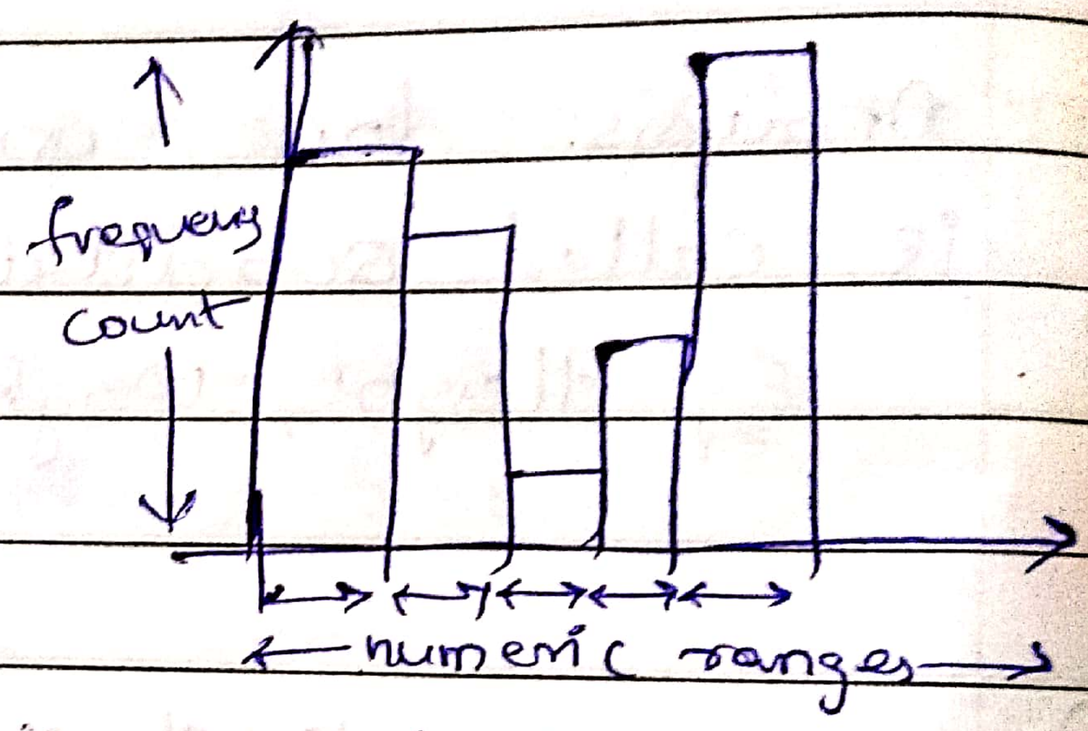
Scores	No. of Students
(65-75)	5
(75-85)	3
(85-95)	1

(7) Barchart & Histogram?

Barchart is the graphical representation of Categorical data.
Histogram is the graphical representation of data, where data is grouped into continuous number ranges.



Barchart



Histogram.

We have gaps in Barchart, as it is defined for categories.
For histogram, we dont have gaps, as it is defined for continuous numerical range.

(8) Five number summary?

Five number summary is a set of descriptive statistics, that provide information about data set. It is also used to find outliers.
It has five steps → min and max values
Q_1, Q_3 and median

Page No.
Date

data set ⇒ {10, 15, 18, 20, 25, 40, 45, 50, 65, 75, 100, 140, 200, 400, 480}

$Q_1 = \left(\frac{n+1}{4}\right)$th term ⇒ $\left(\frac{15+1}{4}\right)$th term ⇒ 4th term

$Q_3 = \frac{3(n+1)}{4}$th term ⇒ 3(4th) ⇒ 12th term

Q_2 = median ⇒ $\left(\frac{n+1}{2}\right)$th term ⇒ 8th term.

∴ $Q_1 = 20$, $Q_2 = 50$, $Q_3 = 75$

$IQR = Q_3 - Q_1 = 75 - 20 = 50$.

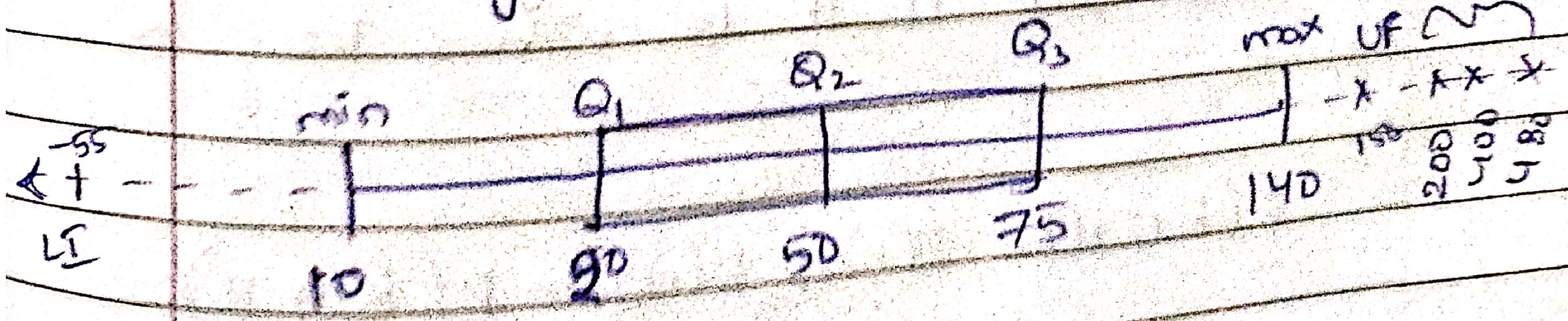
$LI = Q_1 - 1.5(IQR) = 20 - 1.5(50) = -55$

$UI = Q_3 + 1.5(IQR) = 75 + 1.5(50) = 150$

Low inference (LI) = −55

Upper inference (UI) = 150.

∴ (min, max) values from data set from the range (−55, 150) are (10, 140)



Page No.
Date

(9) Addition & multiplication rule?

In probability, we have addition & multiplication rule.

Addition rule:- This is used, when either of the event is going to happen.

- → Mutual exclusive event
- → Non-mutual exclusive event.

Mutual Exclusive event is there is chance of occurance of only one event at a time. For example, if a coin is tossed, we get either side A or side B.

$\therefore P(A \cup B) = P(A) + P(B)$

Non-mutual exclusive event is when two events can occur mutually at a same time. For example, picking a ~~deck~~ card from a deck of cards. Where a card can turn out to be a king from Hearts. To find out if it is either king or Hearts,

$P(K \cup H) = P(K) + P(H) - P(K \cap H)$

Multiplication rule:-

- Independent events
- Dependent event.

Independent events is when two events are occuring independently.
For example, two dices are rolled, and the outcome of 2nd dice is independent on 1st dice (or) vice versa.

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

Dependent event is where 2nd event is dependent on outcome of 1st event.
For example, In a box of 2 red and 3 blue balls, any one ball is removed in 1st event. And the 2nd event is conducted on remaining 4 balls from box which can be either (1 red, 3 blue) or (2 red, 2 blue) balls.

$$P(A \cap B) = P(A) \times P\left(\frac{B}{A}\right)$$

$P(A)$ = 1st event
$P(B/A)$ = 2nd event.

(10) Permutations & combinations :-

Permutation is no. of ways a set can be arranged. Here order of selection in a set is considered

Eg: (A, B, C)
(B, A, C)
(C, A, B)
(C, B, A) etc.

} all are different here

Combination is arrangement of elements in a set, where order is not considered

Eg (A, B, C)
(B, A, C)
(C, B, A)
(C, A, B) etc

} all are same here. what matters is A, B, C are there a set

Example for permutation & combination

How many 3 letters set can be formed from a 11 letter word "Permutation".

$$^nP_r = \frac{n!}{(n-r)!}$$

$n = 11$
$r = 3$.

$$\Rightarrow \frac{11!}{(11-3)!} \Rightarrow \frac{11 \times 10 \times 9 \times 8!}{8!}$$

Page No.
Date

How many combinations of 3 letters set can be formed here =)

$$^{n}C_{r} = \frac{n!}{r!(n-r)!}$$

$$\Rightarrow \frac{11!}{3!\,8!} \Rightarrow \frac{11 \times 10 \times 9 \times \cancel{8!}}{(3 \times 2 \times 1) \times \cancel{8!}}$$

$\Rightarrow 55 \times 3$

$\Rightarrow$ 165 combinations of 3 letters se can be formed here.

(11) Z-Test, T-test, Chi-sq. test.

Z-Test :-

This is a statistical method for the comparision of mean in a sample from the normally distributed population; or between two independent samples.

$$Z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

$\bar{x}$ = mean of sample
μ = mean of population
σ = std. deviation of population
n = no. of samples.

Ex: A teacher claims that mean score of students in class is > 82, with $\sigma = 20$. If a sample of 81 students was selected, with mean score of 90. $\alpha = 0.05$, ~~CF = (-1.96, 1.96)~~. check if there is enough evidence.

$\alpha = 0.05$

$Z_c \Rightarrow 1 - 0.05 \Rightarrow 0.95$.

in z table $Z_c = 0.95$ at 1.6

$\therefore Z = (-1.6, 1.6)$

$H_0 = \mu = 82$

$H_1 \neq \mu > 82$

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$\bar{x} = 90$

$\mu = 82$

$n = 81$

$\sigma = 20$

$$Z = \frac{90 - 82}{20/\sqrt{81}}$$

$$\Rightarrow \frac{\cancel{8}^{2} \times 9}{\cancel{20}_{5}} \Rightarrow Z = \frac{18}{5} = 3.6$$

$Z = 3.6 > 1.6$

$\therefore$ H_0 is rejected

Page No.
Date

T-test :-

This helps us to compare the average values of two data sets and determine, if they belong to same population or different

$$T = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

s = sample std.

$\bar{x}$ = mean of samples
μ = mean of population
σ = std. deviation,
n = no. of samples

one-sample T-test $\Rightarrow \frac{\bar{x} - \mu}{s/\sqrt{n}}$

two-sample T-test $\Rightarrow \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$

Eg: A statistics teacher wants to compare his two classes, to see if their performance.
class A $\Rightarrow$ 25 students, avg score = 70, σ = 15
class B $\Rightarrow$ 20 students, avg score = 74, σ = 25
$\alpha = 0.05$. Did these two classes perform differently

H_0 ; $\mu_{class\ A} = \mu_{class\ B}$
H_1 ; $\mu_{class\ A} \neq \mu_{class\ B}$

* $\alpha = 0.05$

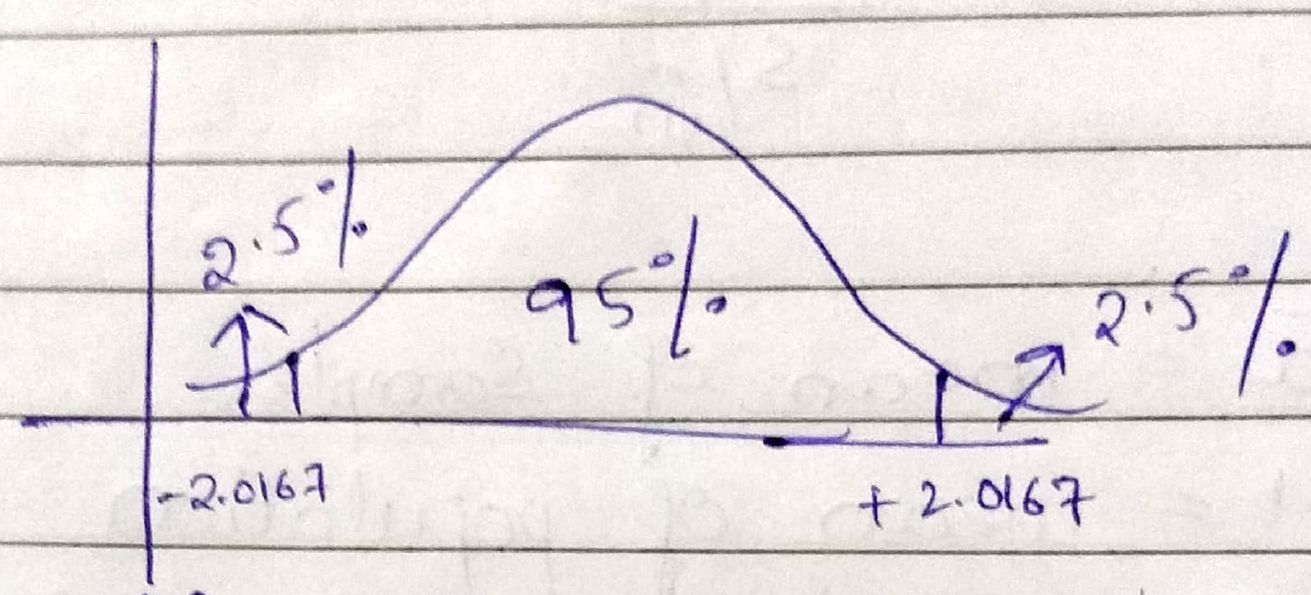
* degree of freedom $\Rightarrow (n_1 - 1) + (n_2 - 1)$

$\Rightarrow (25-1) + (20-1) \Rightarrow 43.$

* State decision rule

$\alpha = 0.05 \Rightarrow 1 - 0.05 \Rightarrow 0.95 \Rightarrow 95\%$

$\Rightarrow$



for $df = 43$, 2 tail test $\alpha = 0.05$

$T = 2.0167.$

$$S_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$$

$SS_1 = S_1^2(df_1) =$

$SS_2 = S_2^2(df_2) =$

$$S_p^2 \Rightarrow \frac{5400 + 11875}{24 + 19} = 401.74$$

$$\therefore T = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} \Rightarrow \frac{70 - 74}{\sqrt{\frac{401.74}{25} + \frac{401.74}{20}}}$$

$\Rightarrow -0.67$

This is in between $(-2.0167, 2.0167)$

$\therefore$ H_0 is accepted.

$\therefore \mu_{class\ A} = \mu_{class\ B}$

Page No.
Date

Chi-square test

This test is used to compare observed results with expected results.

$$\chi^2 = \sum \left\{ \frac{(f_o - f_e)^2}{f_e} \right\}$$

χ^2 = chi-squared value
f_o = observed value
f_e = expected value

Eg

observed

	Republic	Democratic	other	Total
Male	26	13	5	44
female	20	29	7	56
	46	42	12	

} 100

Expected = ?

$$E_{male, Rep} = \frac{(male)(Rep)}{total} = \frac{(44)(46)}{100} = 20.24$$

$$E_{female, Rep} = \frac{(56)(46)}{100} = 25.76$$

$$E_{male, dem} = \frac{(44)(42)}{100} = 18.48$$

$$E_{female, dem} = \frac{(56)(42)}{100} = 23.52$$

$$E_{male,\ other} = \frac{(44)(12)}{100} = 5.28$$

$$E_{female,\ other} = \frac{(56)(12)}{100} = 6.72$$

* H_0 ; party independent on gender
  H_1 ; party dependent on gender

* $\alpha = 0.05$

$$CH = DF - 1$$
$$= (\text{rows} - 1)(\text{columns} - 1)$$
$$= (2-1)(3-1) \Rightarrow \underline{\underline{2}}$$

for $df = 2 \Rightarrow \alpha = 0.05 \Rightarrow 5.991$

$\chi^2 = CH \Rightarrow (-5.991, 5.991)$

* $$\chi^2 = \frac{(26-20.24)^2}{20.24} + \frac{(20-25.76)^2}{25.76} + \frac{(13-13.4}{13.4}$$
$$+ \frac{(29-23.52)^2}{23.52} + \frac{(5-5.28)^2}{5.28} + \frac{(7-6.72)^2}{6.72}$$

$\chi^2 \Rightarrow 5.36$

This is in the range $(-5.991, 5.991)$

$\therefore$ H_0 is accepted

Page No.
Date / /

(12) Binomial & Bernoullis distribution

Binomial distribution is a common discrete distribution used in statistics, as opposed to a continuous distribution, such as normal distribution. This counts only two states, represented as '0' or '1'. 1 for success, 0 for failure.

$$P_x = \binom{n}{x} p^x \cdot q^{n-x}$$

p = binomial probability
x = no. of outcomes
$\binom{n}{x}$ = no. of combinations.
p = Succesful trials
q = failed trials
n = no. of trials

Eg: A dice rolled 10 times. what is the prob of getting 6 for 6 times?

$n = 10$
$x = 6, \; p = 1/6, \; q = 5/6$

$$P = \binom{n}{x} p^x q^{n-x}$$

$$\Rightarrow \frac{10!}{6!\,4!} \left(\frac{1}{6}\right)^6 \left(\frac{5}{6}\right)^4$$

$$\Rightarrow \frac{{}^{5}\cancel{10} \times \cancel{9}^{3} \times \cancel{8}^{2} \times 7 \times \cancel{6!}}{\cancel{6!}} \left(\frac{1 \times 5 \times 5 \times 5 \times 5}{6^{10}}\right)$$

$$\Rightarrow 5 \times 3 \times 2 \times 7 \times \left(\frac{625}{6^{10}}\right)$$

$$\Rightarrow \frac{131250}{60,466,176} \Rightarrow \frac{1}{460.6946}$$

$$p = 0.0023 \Rightarrow 0.23\%$$

$$\text{mean} = n \times p \Rightarrow 10 \times \frac{1}{6} \Rightarrow 1.66$$

$$\text{std.dev} = \sqrt{n \cdot p(1-p)} \Rightarrow \sqrt{10 \times \frac{1}{6} \times \frac{5}{6}}$$

$$\Rightarrow \frac{1}{6}(\sqrt{50}) \Rightarrow 1.178$$

Bernoulli's distribution

This is applied to the events with one trial & two outcomes.

$$P(A) + P(B) = 1$$

Eg: A baby is born probability of the baby to be male is 0.28. Then find the prob of baby to be female.

$$P(M) + P(F) = 1$$
$$0.28 + P(F) = 1$$
$$\therefore P(F) = 1 - 0.28 \Rightarrow 0.72$$