

STATISTICS AND DATA ANALYSIS :-

DATA TYPES:-

Quantitative		Qualitative	
Discrete	Continuous	Nominal	Ordinal
↓	↓	↓	↓
counting whole numbers etc	any value height, weight etc	without order Hair color favourite animal etc	with order Grade (A, B, C) Time of day Financial status

FREQUENCY DISTRIBUTION:-

Frequency Distribution in statistics is a representation of observations within a given interval.

Types of Frequency Distribution:-

1. Grouped Frequency Distribution.
2. Ungrouped Frequency Distribution.

For Example of both Type Sample data:
Marks obtained by 20 Students: 5, 10, 20, 15, 5, 20, 20, 15, 15, 15, 10, 10, 10, 20, 15, 5, 18, 18, 18, 18

Distribution

Grouped Frequency Table:

To arrange a large no of observations or data, we use grouped frequency distribution table. In this we form class intervals to tally the frequency for the data the belongs to that particular class interval.

* Number of classes

$$N = 20 \Rightarrow \text{No of students}$$

$$\therefore m = 1 + 3.3 \cdot \log_{10} N$$

$$m = 1 + 3.3 \cdot \log_{10} 20$$

$$m \approx 5$$

* Range

$$\therefore R = \text{max} - \text{min}$$

$$R = 20 - 5$$

$$R = 15$$

* Class Interval

$$CI = R/m = 15/5 =$$

$$CI = 3$$

Obtained Marks (Class Intervals)	No of Students frequency
5 - 7	3
8 - 10	4
11 - 13	0
14 - 16	5
17 - 19	4
20 - 22	4
Total	20

Ungrouped Frequency Distribution Table.

In Ungroup Frequency table we dont make classes Intervals. we write the accurate frequency of Individual data.
 Or you can say we write frequency of a category.

Marks Obtained	No Students
5	3
10	4
15	5
18	4
20	4
Total	20

LAB I

- a) Collect the data set in a tabular form: Done
- b) Construct a frequency distribution of height/weight data
- c) Using frequency distribution as construct draw a histogram and superimpose it on a frequency polygon and frequency curve
- d) write comment on the data distribution

Solving

b)

For Height:

* Number of Classes

$$\therefore N = 67$$

$$\therefore m = 1 + 3 \cdot 3 \log_{10} N$$

$$m \approx 7$$

* Range

$$R = \text{max} - \text{min} = 191 - 145 = 46$$

$$R = 46$$

* Class Interval

$$CI = \frac{\text{Range}}{\text{Class size}} = R/m = 46/7 = 7$$

$$CI = 7$$

d)

Classes

Class
Boundaries

Tally
Mark

Mid
Point

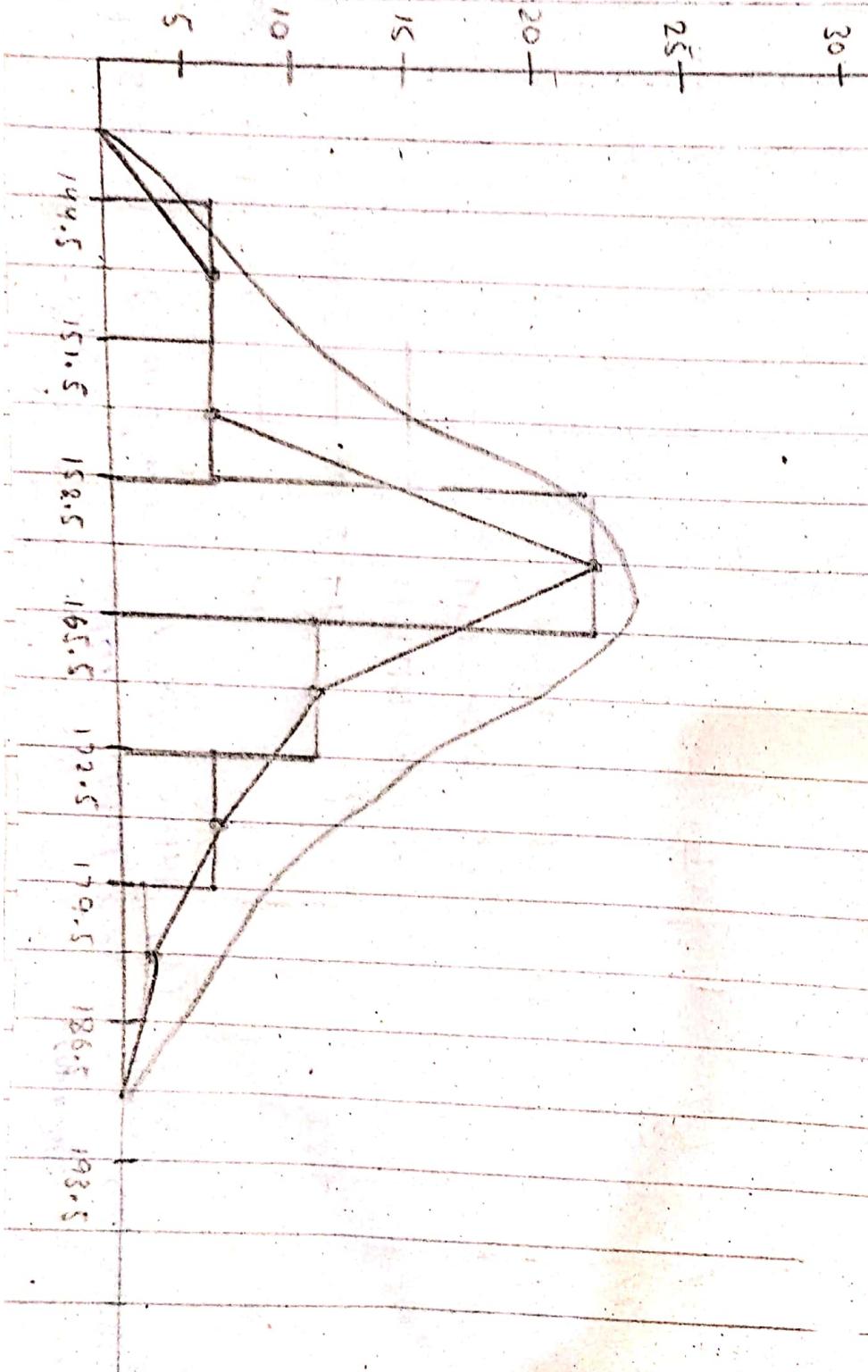
frequency
(f)

frequency
(cf)

Probability
(P)

145 – 151	144.5 – 151.5		148	6	6	6/67
152 – 158	151.5 – 158.5		155	6	6	6/67
159 – 165	158.5 – 165.5		162	15	15	15/67
166 – 172	165.5 – 172.5		169	22	27	22/67
173 – 179	172.5 – 179.5		176	11	60	11/67
180 – 186	179.5 – 186.5		183	6	66	6/67
187 – 193	186.5 – 193.5		190	1	67	1/67

Q c)



LAB II: DATA SUMMARIZATION:-

Problem: Consider the Student height/weight data from table 1, and Summarize the Information.

↓ ↓ ↓
Mean Median Mode.

DATA Type/ Statistics	MEAN	MEDIAN	MODE
Ungrouped	$\bar{x} = \frac{\sum u}{n}$	$\bar{x} = \left(\frac{n+1}{2} \right)^{th}$ when n is odd when n is even $\bar{x} = \left\{ \left(\frac{n}{2} \right) + \left(\frac{n+2}{2} \right) \right\} / 2$	$\bar{x} = \left(\frac{n+1}{2} \right)^{th}$ Mode is most frequent value
Grouped	$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$	$\bar{x} = x_L + \frac{\left(\frac{n}{2} - c \right) * h}{f}$	Mode = $x_L + \frac{f_m - f_i}{2f_m - f_i - f_2} * h$

Variables:-

$\sum u$ = Sum of all Students height

n = Total no of Students = $\sum f$

In Group data

MEDIAN Variables:-

x_l = MEDIAN Class lower boundary

f = MEDIAN Class frequency

h = Class Interval

C = MEDIAN Class se piche walli

class kei Commulative frequency

MODE Variables:-

f_m = MEDIAN Class frequency

f_1 = MEDIAN Class before frequency

f_2 = MEDIAN Class after frequency

LAB III.

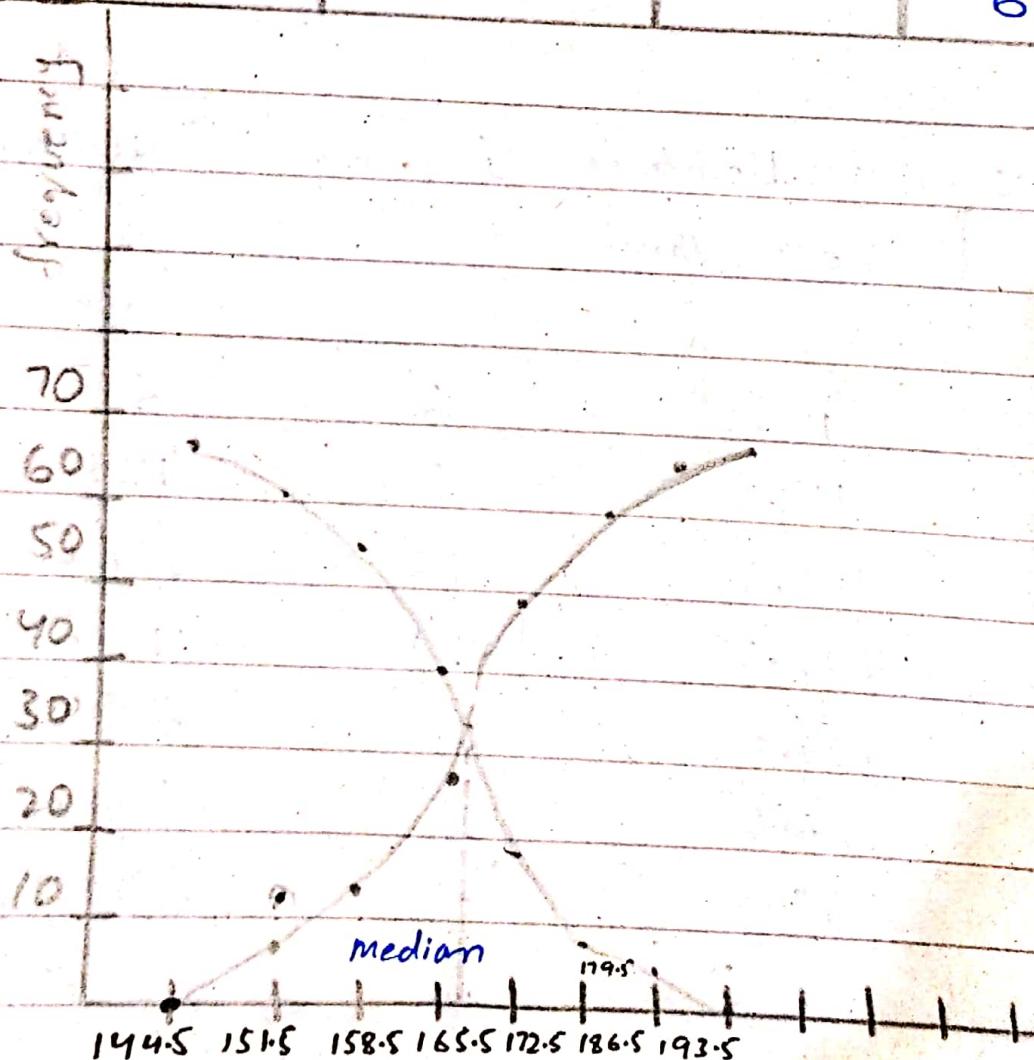
Problem: Consider the frequency distribution of Height/weight Data in Lab 1

- a) Obtain the Median graphically
- b) Obtain Mode graphically
- c) compare values numerically with values completed in lab 2.

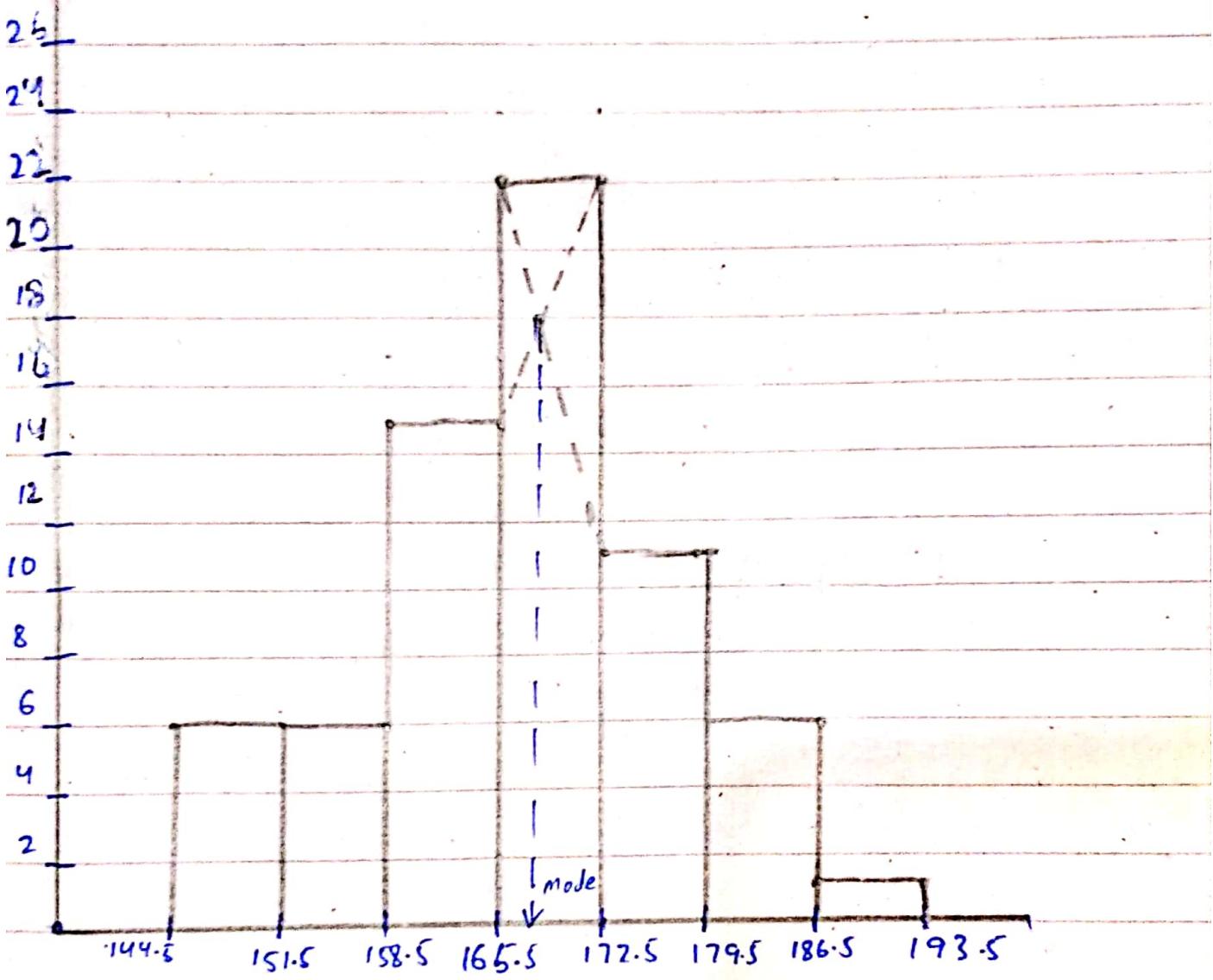
a)

Class Boundaries	Class Boundaries Less than	frequency	frequency less than Type
149.5 — 151.5	151.5	6	6
151.5 — 158.5	158.5	6	12
158.5 — 165.5	165.5	15	27
165.5 — 172.5	172.5	22	49
172.5 — 179.5	179.5	11	60
179.5 — 186.5	186.5	6	66
186.5 — 193.5	193.5	1	67

C.B	C.B More than	f	f more than
144.5 — 151.5	144.5	6	67
151.5 — 158.5	151.5	6	61
158.5 — 165.5	158.5	15	55
165.5 — 172.5	165.5	22	40
172.5 — 179.5	172.5	11	18
179.5 — 186.5	179.5	6	7
186.5 — 193.5	186.5	1	6



$$\text{Median} = 167.5$$



$$\text{Mode} = 167.8$$

	MEDIA	MODE
From	167.014	168.2
<u>LAB II</u>		

From	167	167.8
<u>LAB III</u>		

LAB IV Shape of Distribution

Problem:

Consider the height/weight data collected in practical # 1

- a) calculate all measure of dispersion from grouped and ungrouped data set and write comments
- b) consider the grouped data of height and compute - Skewness and kurtosis during first four moments.
- c) consider the ungrouped data of height calculate first four moments and calculate Skewness and kurtosis.

a)

For Group Data:

Range Deviation:-

$$R.D = \frac{Max - Min}{2}$$

$$R.D = \frac{191 - 145}{2}$$

$$R.D = 23$$

coefficient of Range

$$CRD = \frac{Max - Min}{Max + Min}$$

$$CRD = \frac{191 - 145}{191 + 145}$$

$$CRD = 0.136$$

Quartile Deviations:-

$$\therefore Q.D = \frac{Q_3 - Q_1}{2}$$

Jts for
ngroup data

$$Q_i = L + \left(\frac{\frac{n}{4} - C}{f} \right) h$$

For group
data

Max Mid point

and Min Mid
point

$$\Rightarrow Q_1 = L + \left(\frac{\frac{n}{4} - C}{f} \right) \times h$$

$$\text{for } Q_1 \quad n/4 = 67/4$$

$$\frac{67}{4} \rightarrow 16.75$$

Class Boundaries	f	c.f	
144.5 — 151.5	6	6	
151.5 — 158.5	6	12	
158.5 — 165.5	15	27	← Q ₁
165.5 — 172.5	22	49	
172.5 — 179.5	11	60	← Q ₃
179.5 — 186.5	6	66	
186.5 — 193.5	1	67	

$$Q_1 = L + \left(\frac{3n/4 - C}{f} \right) h$$

$$\boxed{Q_1 = 160.716}$$

Similarly for Q₃

$$Q_3 = L + \left(\frac{\frac{3n}{4} - C}{f} \right) h$$

$$\frac{3n}{4} \rightarrow \frac{3(67)}{4} \rightarrow 50.25$$

$$Q_3 = 172.5 + \left(\frac{50.25 - 49}{11} \right) \times 7$$

$$\boxed{Q_3 = 173.2954}$$

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{173.2954 - 160.716}{2}$$

$$Q.D = 6.2897$$

Coefficient of QD

MEAN DEVIATION:-

$$C.Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = 3.7 \times 10^{-6}$$

$$Q_3 + Q_1$$

$$\text{Deviation} = \text{Midpoint} - C$$

Mean ↓
 Median Mode

$$\therefore M.D = \frac{\sum f |x_i - \bar{x}|}{\sum f}$$

Mean deviation table.

Classes	f	Mid point x_i	fx	$ x_i - \bar{x} $	$f x_i - \bar{x} $
145 - 151	6	148	888	19.014	114.084
152 - 158	6	155	930	12.014	72.084
159 - 165	15	162	2430	5.014	75.21
166 - 172	22	169	3718	1.986	43.692
173 - 179	11	176	1936	8.986	98.846
180 - 186	6	183	1098	15.986	95.916
189 - 193	1	190	190	22.986	22.986
	<u>67</u>		<u>11190</u>	<u>85.986</u>	<u>522.818</u>

$$\text{Mean} = \bar{x} = \frac{\sum f x_i}{n} = \frac{11190}{67}$$

$$\boxed{\bar{x} = 167.014}$$

$$M.D = \frac{\sum f|x_i - \bar{x}|}{\sum f} = \frac{522.818}{67}$$

$$\boxed{M.D = 7.803}$$

Standard Deviation:-

$$S.D = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum D^2}{n}}$$

$$D = (x_i - \bar{x}) \rightarrow \text{Mid point} - \text{Mean}$$

$$S.D = \sqrt{\frac{\sum FD^2}{\sum F} - \left(\frac{\sum FD}{\sum F}\right)^2}$$

Variance:-

$$\sigma^2 = (S.D)^2$$

$$\sigma^2 = \frac{\sum FD^2}{\sum F} - \left(\frac{\sum FD}{\sum F}\right)^2$$

Coefficient of S.D

$$CSD = \frac{S.D}{\bar{x}} \rightarrow \text{mean}$$

Coefficient of variations:-

$$CV = \frac{S.D}{\bar{x}} * 100$$

Range Deviations:-

$$R.D = \frac{\text{Max Mid point} - \text{Min Mid point}}{2}$$

↓
Grouped data.

Ungroup Data:-

Quartile Deviations:-

$$Q.D = \frac{Q_3 - Q_1}{2}$$

$$Q_i = \left(\frac{i n}{4} \right)^{\text{th}} \text{ Orderd data.}$$

MEAN DEVIATION:-

$$MD = \frac{\sum |x_i - C|}{n}$$

C
→ Median
→ Mean
→ Mode

Standard Deviation:-

$$S.D = \sqrt{\frac{\sum D^2}{n} - \left(\frac{\sum D}{n} \right)^2}$$

$$D = (x_i - \bar{x}) = \text{individual data} - \text{Mean.}$$

Variance:-

$$\sigma^2 = (S.D)^2$$

$$\sigma^2 = \frac{\sum D^2}{n} - \left(\frac{\sum D}{n} \right)^2$$

ABSOLUTE MEASURE OF Dispersion.

Range Deviation:-

~~Excess~~

Problem: Only two observations.

Quartile Deviation:-

Problem:

- * It does not depends on all values
- * It affected by only two observations

Definition:-

It is also called semi-inter quartile range. It is the half of the product of the upper and lower quartile or $\frac{P_{75} - P_{25}}{2}$.

MEAN DEVIATION:-

It is defined as the average of the absolute deviation taken from some central value such as (Mean, Median and Mode)

Points.

- based on all values.
- Simple to calculate.
- not effected by extreme values
- Not accurate Measure of dispersion
- It is not widely used.

Standard Deviation:-

It is defined as the positive square root of the average of the squared deviations, taken from A.M of the distribution Mathematically.

S.D Formulas.

Ungroup Data

$$S.D = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

Grouped Data

$$S.D = \sqrt{\frac{\sum f_i D^2}{\sum f_i}}$$

$$S.D = \sqrt{\frac{1}{n} \left[\sum D_i^2 - \left(\frac{\sum D_i}{n} \right)^2 \right]}$$

$$S.D = \sqrt{\frac{\sum f_i D^2 - \left(\frac{\sum f_i D}{\sum f_i} \right)^2}{\sum f_i}}$$

Standard deviation is independent of change of origin and scale.

ROOT MEAN SQUARED DEVIATION:-

It is defined as positive square root of average of deviation taken from some arbitrary constant (A)

$$S = \sqrt{\frac{\sum f_i (x_i - A)^2}{n}}$$

Relation b/w R.M.S.D and S.D

$$S^2 = \sigma^2 + (\bar{x} - A)^2$$

if $\bar{x} = A$ then

$$\begin{matrix} S^2 &= \sigma^2 \\ \downarrow & \downarrow \\ \text{R.M.S.D} & \text{S.D} \end{matrix}$$

Points.

- It may be manipulated Mathematically.
- It is best measure of dispersion
- It is based on all the observations
- It is of great importance of sampling theory
- It is least effected by Sampling fluctuations
- It is not good measure when the series of data is short.

Variance:

Variance is not an independent measure of dispersion it is defined as the squared of Standard deviation.

Moments:-

moments of the distribution is designated by the power of the deviation, varies to at end, when deviation is taken about mean then moments are called as mean moments whereas the deviation about any constant A (when $A=0$) then moments are called raw moments or moments about origin. and order of moments designated by the power of deviation.

Raw Moments:-

when $A=0$

Group

$$\therefore M_r = \frac{1}{\sum f_i} \sum f_i x_i^r$$

ungroup

$$\therefore M_r = \frac{1}{n} \sum x_i^r$$

x_i = Mid point . ~~Deviation~~ x_i = Individual observation

MEAN MOMENTS:-

Group

$$M_r = \frac{1}{\sum f_i} \sum f_i (x_i - \bar{x})^r$$

ungroup

$$M_r = \frac{1}{n} \sum (x_i - \bar{x})^r$$

Points

Moments are independent of change of origin but dependent of change of scale.

Relation b/w Raw Moments and Mean Moments:-

Raw Moments about Origin.

$$M'_1 = \frac{1}{n} \cdot \sum x_i = \text{Mean} = \bar{x}$$

$$M'_2 = \frac{1}{n} \cdot \sum x_i^2$$

$$M'_3 = \frac{1}{n} \cdot \sum x_i^3$$

$$M'_4 = \frac{1}{n} \sum x_i^4$$

Moments about Mean

For M_1 ,

$$M_1 = \frac{1}{n} \cdot \sum (x_i - \bar{x})$$

mean also mean

$$U_1 = \frac{1}{n} \cdot \mathbb{E}(x_i) - \frac{1}{n} \mathbb{E}(\bar{x})$$

$U_1 = 0$ Proof

For U_2

$$U_2 = \frac{1}{n} \cdot \mathbb{E}(x_i - \bar{x})^2$$

$U_2 = \frac{1}{n} \mathbb{E}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$ he hogya.

$U'_2 = \frac{1}{n} \cdot \mathbb{E}x_i^2 - 2\bar{x} \cdot \mathbb{E}x_i + \bar{x}^2$

$$U_2 = U'_2 - 2U'_1 \cdot U'_1 + U'^2_1$$

$$\boxed{U_2 = U'_2 - U'^2_1}$$

For U_3

$$U_3 = \frac{1}{n} \cdot \mathbb{E}(x_i - \bar{x})^3$$

$$U_3 = \frac{1}{n} \mathbb{E}(x_i^3 - 3x_i^2\bar{x} + 3x_i\bar{x}^2 - \bar{x}^3)$$

$$U_3 = \frac{1}{n} \mathbb{E}x_i^3 - 3\bar{x} \frac{\mathbb{E}x_i^2}{n} + 3\bar{x} \frac{\mathbb{E}x_i}{n} - \bar{x}^3$$

\bar{x} pe summision
lag kee n
se devide ho
to ans mean

$$M_3 = M'_3 - 3M'_1 M'_2 + 3M'_1 \cdot M'_1 - M'^3$$

$$\boxed{M_3 = M'_3 - 3M'_1 M'_2 + 2M'^2_1}$$

For M_4

$$M_4 = \frac{1}{n} \cdot \sum (x_i - \bar{x})^4$$

$$M_4 = \frac{1}{n} \cdot \sum (x_i^4 - 4x_i^3 \bar{x} + 6x_i^2 \bar{x}^2 - 4x_i \bar{x}^3 + \bar{x}^4)$$

$$M_4 = \frac{1}{n} \cdot \sum x_i^4 - 4\bar{x} \frac{\sum x_i^3}{n} + 6\bar{x}^2 \frac{\sum x_i^2}{n} - 4\bar{x}^3 \frac{\sum x_i}{n} + \bar{x}^4$$

$$M_4 = M'_4 - 4M'_1 M'_3 + 6M'^2_1 M'_2 - 4M'^3_1 M'_1 + M'^4_1$$

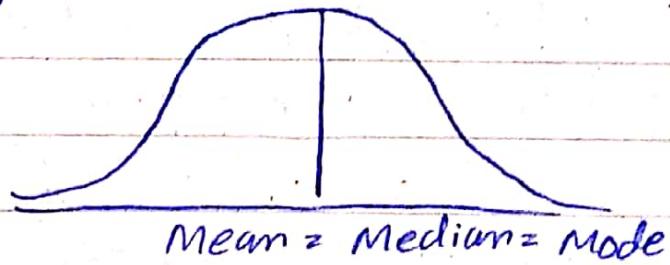
$$M_4 = M'_4 - 4M'_1 M'_3 + 6M'^2_1 M'_2 - 4M'^3_1 + M'^4_1$$

$$\boxed{M_4 = M'_4 - 4M'_1 M'_3 + 6M'^2_1 M'_2 - 3M'^4_1}$$

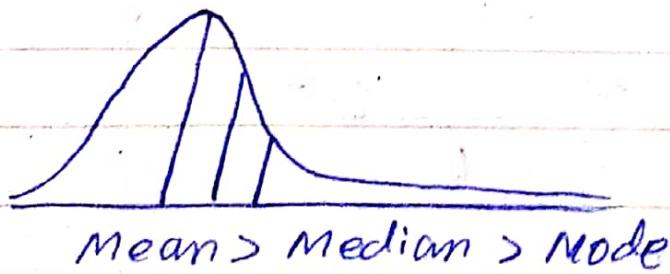
Skewness-

The degree of departure of frequency curve of a distribution from symmetric nature is called Skewness of the distribution.

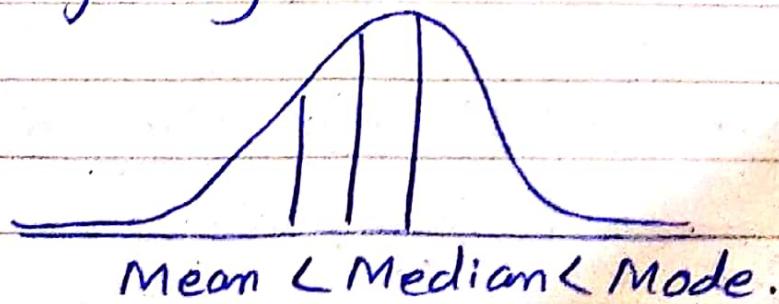
1 Symmetric \Rightarrow Mean = Median = Mode



2 - Positively skewed \Rightarrow Mean > Median > Mode



3- Negatively skewed



MEASURE OF SKEWNESS:-

$$\beta_1 = \frac{\mu_2^2}{\mu_1^3}$$

β_1 is absolute measure of skewness

If

$\beta_1 = 0$ distribution is symmetric

$\beta_1 > 0$ distribution is skewed either positively or negatively.

Relative Measure of Skewness.

Acc to Karl Pearson.

$$Sk = \frac{\text{Mean} - \text{Mode}}{S.D}$$

Using relation b/w mean, Median and Mode. (Empirical relation)

Hence

$$S.K = \frac{3(\text{Mean} - \text{Median})}{S.D}$$

also

$$S.K = \frac{3(\text{Mean} - Q_2)}{S.D}$$

According to Bowley.

$$S.I.C = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

$$0 \leq \text{Skewness} \leq +1$$

3.

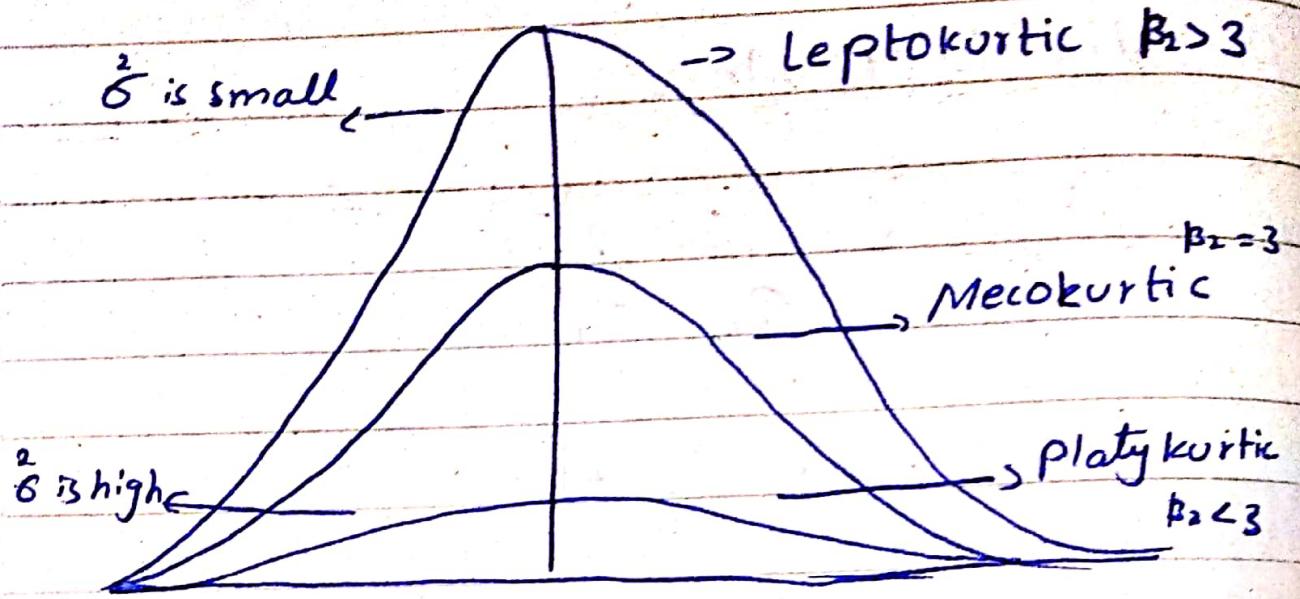
$$\text{Skewness} = \frac{\sqrt{B_1} (B_2 + 3)}{2 [5B_2 - 6B_1 - 9]}$$

MEASURE OF kurtosis:-

B_2 is measure of kurtosis

$$B_2 = \frac{\mu_4}{\mu_2^2}$$

* It is a characteristic of frequency distribution. It measures flatness and peakness of curve of the given distribution.



* Show that $B_2 \geq 1$

$$B_2 = \frac{\mu_4}{\mu_2^2}$$

$$\frac{\mu_4}{\mu_2^2} \geq 1$$

$$\mu_4 \geq \mu_2^2$$

$$\therefore \mu_4 = \frac{\sum f(x_i - \bar{x})^4}{\sum f}$$

$$\therefore \mu_2 = \frac{\sum f(x_i - \bar{x})^2}{\sum f}$$

$$\frac{\sum f(x_i - \bar{x})^4}{\sum f} \geq \left[\frac{\sum f(x_i - \bar{x})^2}{\sum f} \right]^2$$

Let
Let $d = (x_i - \bar{x})^2$

$$\frac{\sum f d^2}{\sum f} \geq \left(\frac{\sum f d}{\sum f} \right)^2$$

$$\frac{\sum f d^2}{\sum f} - \left(\frac{\sum f d}{\sum f} \right)^2 \geq 0$$

$$\boxed{\sigma^2 \geq 0}$$

It is always true by using
the property of variance

By 4 $\boxed{\beta \geq 1}$

Theorem:- Show that coefficient of kurtosis
is greater than equals to coefficient
of skewness.

$$\cancel{\star} : \beta_2 \geq \beta_1$$

$$\beta_1 = \frac{m_3^2}{m_2^3}, \quad \beta_2 = \frac{m_4}{m_2^2}$$

$$\therefore \beta_2 \geq \beta_1$$

$$\rightarrow \frac{m_4}{m_2^2} \geq \frac{m_3^2}{m_2^3}$$

$$\rightarrow m_2 m_4 \geq m_3^2 \quad \text{--- (1)}$$

$$\because m_2 = \frac{\sum f d^2}{n} \quad \left. \right\}$$

$$\because m_3 = \frac{\sum f d^3}{n} \quad \left. \right\} \rightarrow d = (x; -\bar{x})$$

$$\therefore m_4 = \frac{\sum f d^4}{n}$$

(1) \Rightarrow

$$\left[\frac{\sum f d^2}{n} \right] \left[\frac{\sum f d^4}{n} \right] \geq \left[\frac{\sum f d^3}{n} \right]^2$$

$$\varepsilon f d^2 \cdot \varepsilon f d^4 \geq (\varepsilon f d^3)^2$$

LAB V

Box Plot / Stem & Leaf Plot:-

Stem & Leaf plot is a method for presenting quantitative data in a graphical format, similar to a histogram, to assist in visualising the shape of distribution.

Example data:-

1.17	1.61	1.16	1.38	3.53
1.23	3.76	1.94	0.96	4.75
0.15	2.41	0.71	0.02	1.59
0.19	0.82	0.47	2.16	2.01
0.92	0.75	2.59	3.07	1.40

STEMS	LEAFS
0	15 19 82 71 47 96 02 92 75
1	17 23 61 16 94 38 59 40
2	41 59 16 01
3	76 07 53
4	75

STEMS	LEAVES	FREQUENCY	PROBABILITY
5	0	1	0.04
5	1	2	0.08
5	2	1	0.04
5	3	1	0.04
5	4	1	0.04
5	5	1	0.04
5	6	1	0.04
5	7	1	0.04
5	8	1	0.04
5	9	1	0.04
6	0	1	0.04
6	1	1	0.04
6	2	1	0.04
6	3	1	0.04
6	4	1	0.04
6	5	1	0.04
6	6	1	0.04
6	7	1	0.04
6	8	1	0.04
6	9	1	0.04
7	0	1	0.04
7	1	1	0.04
7	2	1	0.04
7	3	1	0.04
7	4	1	0.04
7	5	1	0.04
7	6	1	0.04
7	7	1	0.04
7	8	1	0.04
7	9	1	0.04
8	0	1	0.04
8	1	1	0.04
8	2	1	0.04
8	3	1	0.04
8	4	1	0.04
8	5	1	0.04
8	6	1	0.04
8	7	1	0.04
8	8	1	0.04
8	9	1	0.04
9	0	1	0.04
9	1	1	0.04
9	2	1	0.04
9	3	1	0.04
9	4	1	0.04
9	5	1	0.04
9	6	1	0.04
9	7	1	0.04
9	8	1	0.04
9	9	1	0.04
		$n=25$	
		1	

$$Q_i = \left(\frac{\text{in}}{4}\right)^{\text{th}}$$

$$Q_1 = \frac{25}{4} = (18)^{\text{th}} = \cancel{0.82} \quad 0.82$$

$$Q_2 = \frac{2(25)}{4} = (13)^{\text{th}} = 1.38$$

$$Q_3 = \frac{3(25)}{4} = (19)^{\text{th}} = 2.16$$

$$Q_4 = \frac{4(25)}{4} = (25)^{\text{th}} = 4.75$$

$$\text{IQR} = Q_3 - Q_1 = 2.16 - 0.82$$

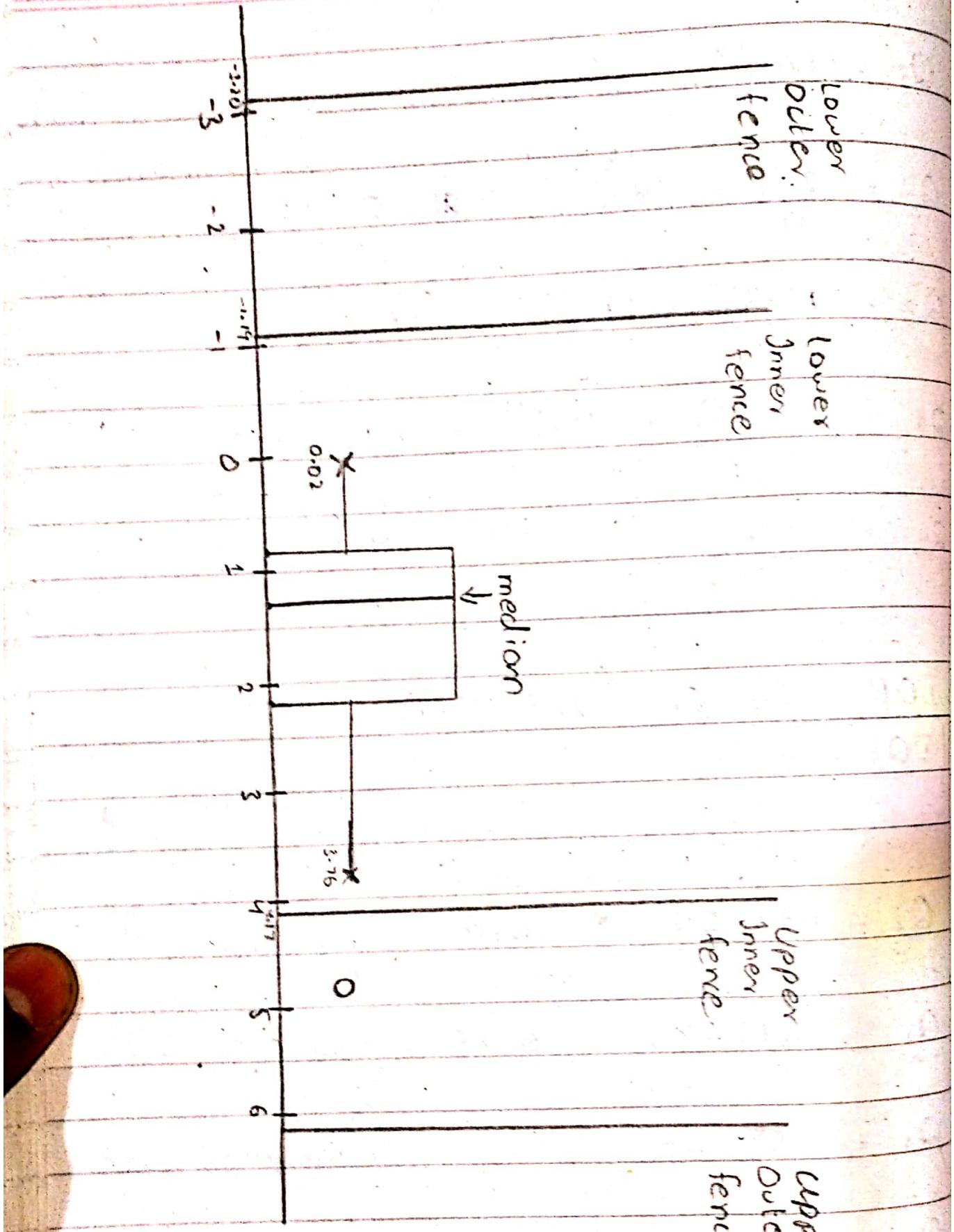
$$\boxed{\text{IQR} = 1.34}$$

$$Q_2 - 1.5(\text{IQR}) = -1.19 \rightarrow \text{Lower Inner fence}$$

$$Q_1 - 3(\text{IQR}) = -3.20 \rightarrow \text{Lower Outer fence}$$

$$Q_3 + 1.5(\text{IQR}) = 4.17 \rightarrow \text{Upper inner fence}$$

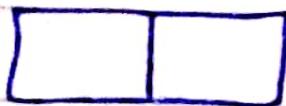
$$Q_3 + 3(\text{IQR}) = 6.18 \rightarrow \text{Upper outer fence.}$$



Hence middle line of box is near to Quartile 1 so we can say the distribution is Positively skewed.



→ Positively skewed.



→ Symmetric



→ Negatively skewed.

Probability:-

A phenomenon is called random if the exact outcome is uncertain, the mathematical ~~outcome~~ study of such randomness is called the theory of probability.

Event:-

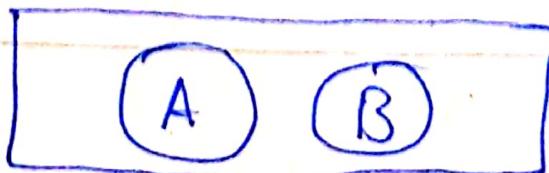
A probability event can be defined as a set of outcomes of an experiment, in other words, an event in probability is the subset of the respective sample space.

Sample Space:-

All possible outcomes collection.

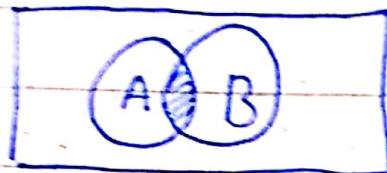
Mutually Exclusive Events:-

Two events are mutually exclusive when no elements are common between them.



Non - Mutually Exclusive Events:-

Two Events are Non-Mutually Exclusive
If they Share common Elements.



VEN Diagrams:-

A Ven diagram is a diagram which related in set theory in mathematics by which happening of events can be visualize.

Independent Events:-

If the occurrence of events cannot effect occurrence of another event

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2)$$

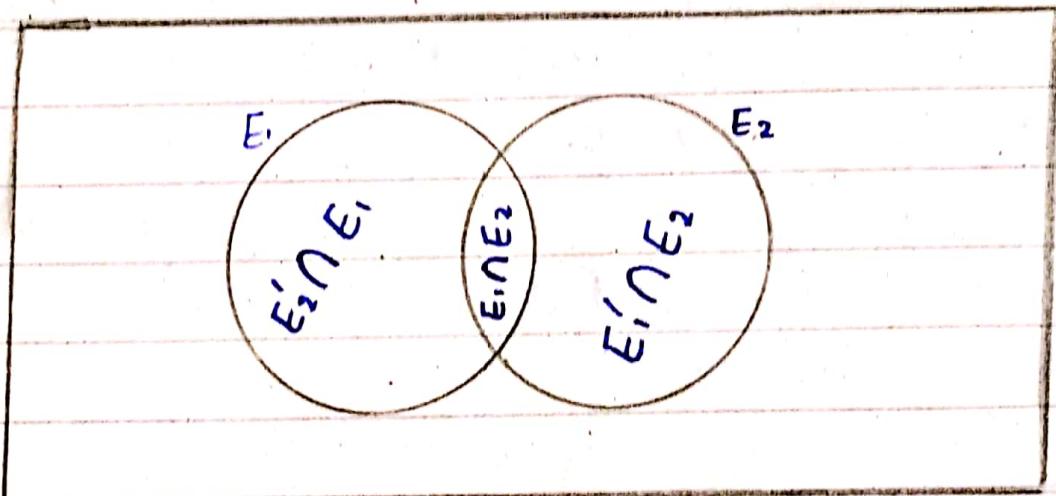
$$P(E_1 \cap E_2 \cap E_3) = P(E_1) \cdot P(E_2) \cdot P(E_3)$$

Theorem:-

If E_1 & E_2 be two events define on the sample spaces, then the probability that E_1 occurs or E_2 occurs or both of them occurs is given by.

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Proof:



$$E = (E_1 \cup E_2) = E_1 \cup (E_1' \cap E_2)$$

$$P(E) = P(E_1 \cup E_2) = P[E_1 \cup (E_1' \cap E_2)]$$

$$= P(E_1 \cup E_2) = P(E_1) + P(E_1' \cap E_2)$$

→ Adding and Subtracting $P(E_1 \cap E_2)$

$$= P(E_1 \cup E_2) = P(E_1) + \underline{P(E_1' \cap E_2)} + P(E_1 \cap E_2) - P(E_1 \cap E_2)$$

$$P(E) = P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Theorem #2 :-

If E_1, E_2 & E_3 be the events define on the sample spaces, then the probability that E_1 occur or E_2 or E_3 or at least one of them occur is given by.

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3) &= P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) \\ &\quad - P(E_1 \cap E_3) - P(E_2 \cap E_3) + \\ &\quad P(E_1 \cap E_2 \cap E_3) \end{aligned}$$

Proof:

Here E and E_3 be two Events define on S , then.

$$P(E \cup E_3) = P(E) + P(E_3) - P(E \cap E_3)$$

In above theorem

$$\therefore P(E) = P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

$$P(E_1 \cup E_2 \cup E_3) = P(E_1 \cup E_2) + P(E_3) - P((E_1 \cup E_2) \cap E_3)$$

$$\therefore [(E_1 \cup E_2) \cap E_3] = (E_1 \cap E_3) \cup (E_2 \cap E_3) \text{ distributive}$$

$$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P[(E_1 \cap E_3) \cup (E_2 \cap E_3)]$$

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3) &= P(E_1) + P(E_2) + P(E_3) - [P(E_1 \cap E_3) + \\ &\quad P(E_2 \cap E_3) - P(E_1 \cap E_2 \cap E_3)] \end{aligned}$$

Theorem

$$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_3) \\ - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3)$$

$$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - \underline{P(E_1 \cap E_2)} \\ - P(E_1 \cap E_3) - P(E_2 \cap E_3) + \\ P(E_1 \cap E_2 \cap E_3)$$

missed it
in above
steps.

For General Formula:-

$$S_1 = P(E_1) + P(E_2) + \dots + P(E_n)$$

$$S_2 = P(E_i \cap E_j) ; \text{ for } i < j$$

$$S_3 = P(E_i \cap E_j \cap E_k) ; \text{ for } i < j < k$$

.

.

$$S_n = P(E_1 \cap E_2 \cap \dots \cap E_n)$$

$$P(E_1 \cup E_2 \cup \dots \cup E_n) = S_1 - S_2 + S_3 + \dots + (-1)^{n+1} S_n$$

LAB VI : Probability:-

Problems:-

Two Dices are rolled

- Construct the sample space
- Let n be the random variable that denotes the sum of the scores on the Two dice
- Construct the probability distribution for the random variable x
- repeat it for four sided dice.

01 02	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$$P(n) = \frac{n(E)}{n(S)}$$

Probability Distribution.

n	2	3	4	5	6	7	8	9	10	11	12
$P(n)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

LAB VI : Probability :-

Problems:-

Two Dices are rolled

- Construct the sample space
- Let n be the random variable that denotes the sum of the scores on the Two dice
- Construct the probability distribution for the random variable x
- repeat it for four sided dice.

O_1	1	2	3	4	5	6	
O_2	1	2	3	4	5	6	7
1	2	3	4	5	6	7	8
2	3	4	5	6	7	8	9
3	4	5	6	7	8	9	10
4	5	6	7	8	9	10	11
5	6	7	8	9	10	11	12
6	7	8	9	10	11	12	

$$P(n) = \frac{n(E)}{n(S)}$$

Probability Distribution.

n	2	3	4	5	6	7	8	9	10	11	12
$P(n)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Random Variables:-

It is a real valued function which assign a real number to each sample point in the sample space.

Example :-

Three Coins

$$S = \{ HHH, HHT, HTH, THH, HTT, THT, TTH, TTT \}$$

for Heads

$$X(S_1) = 3$$

$$X(S_2) = X(S_3) = X(S_4) = 2$$

$$X(S_5) = X(S_6) = X(S_7) = 1$$

$$X(S_8) = 0$$

Random variable $n \in \mathbb{N} \& n \geq 0$

Discrete Probability Distribution Function.

$X(\text{no of heads})$	0	1	2	3	T
$P(n)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

Probability Mass functions:- P.M.F

Following Conditions:-

$$\textcircled{1} \quad P(u_i) \geq 0$$

$$\textcircled{2} \quad \sum P(u_i) = 1$$

Distribution Functions:-

It is define as cumulative distribution function.

$$f(n) = \sum P_i$$

$$f(n) = \begin{cases} 1/8 & n \leq 0 \\ 3/8 & n \leq 1 \\ 7/8 & n \leq 2 \\ 1 & n \leq 3 \end{cases}$$

Axioms in Probability:-

- i) $P(\emptyset) = 0$; If an event is certain not to occur its probability is zero.
- ii) $P(s) = 1$; If an event is certain to occur its probability is unity.
- iii) $0 \leq P(E) \leq 1$; from (i) and (ii)
- iv) $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$
Non Mutually Exclusive Events.

PROBLEM:-

The Probability that a person visiting his dentist will have his teeth cleaned, or a cavity filled, a tooth extracted, his teeth cleaned and a cavity filled, his teeth cleaned and a tooth extracted, a cavity filled and a tooth extracted, or his teeth cleaned and cavity filled and a tooth extracted are 0.47, 0.29, 0.22, 0.08, 0.06, 0.07 and 0.03 what is the probability that a person will have at least one of them is done?

DATA:-

E_1 = Tooth Cleaned

E_2 = Cavity filled

E_3 = Tooth extracted

$$P(E_1) = 0.47$$

$$P(E_2) = 0.29$$

$$P(E_3) = 0.22$$

$$P(E_1 \cap E_2) = 0.08$$

$$P(E_1 \cap E_3) = 0.06$$

$$P(E_2 \cap E_3) = 0.07$$

$$P(E_1 \cap E_2 \cap E_3) = 0.03$$

$$P(E_1 \cup E_2 \cup E_3) = ?$$

we know that.

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3) &= P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) \\ &\quad - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3) \end{aligned}$$

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3) &= 0.47 + 0.29 + 0.22 - 0.08 - \\ &\quad 0.06 - 0.07 + 0.03 \end{aligned}$$

$$P(E_1 \cup E_2 \cup E_3) = 0.8$$

Difference between Mutually exclusive and Independent events:-

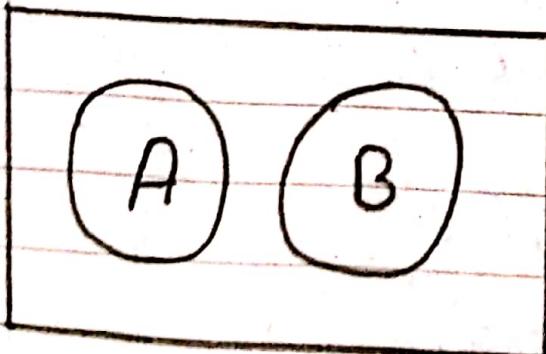
Mutually Exclusive EVENTS:-

when the occurrence is not simultaneous for two events then they are termed as Mutually Exclusive events

The non-occurrence of an event will end up in the occurrence of another event.

Formula:

$$P(E_1 \cap E_2) = 0$$



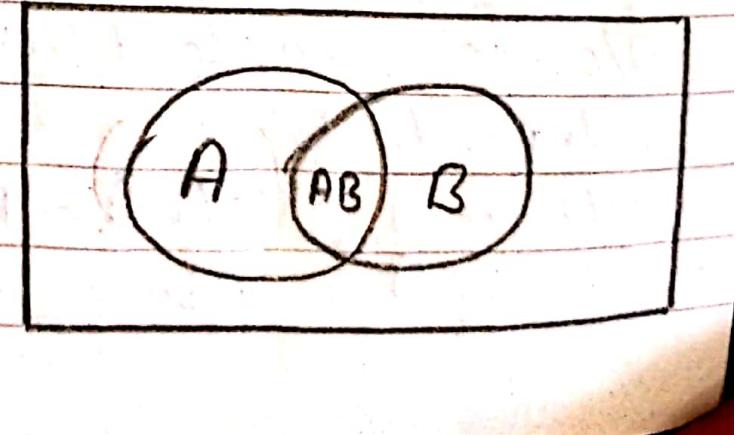
INDEPENDENT EVENTS:-

when the occurrence of one event does not control the happening of the other event the it is termed as an independent event

There is no influence of an occurrence with another and they are independent of each other.

Formula:

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2)$$



Theorem 1:-

If E_1 & E_2 are Independent then
Show that E_1 and E_2' are also
Independent:-

Proof:

Given:

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2)$$

To Prove:

$$P(E_1 \cap E_2') = P(E_1) \cdot P(E_2')$$

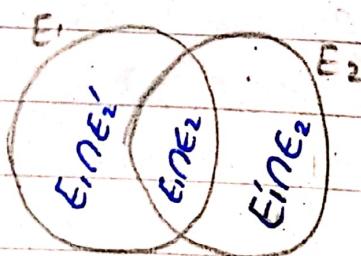


fig: 1

A/c to figure 1

$$\rightarrow P(E_1 \cap E_2') = P(E_1) - P(E_1 \cap E_2)$$

A/c to given condition that E_1 & E_2 are
Independent

$$\rightarrow P(E_1 \cap E_2') = P(E_1) - P(E_1) P(E_2)$$

$$\rightarrow P(E_1 \cap E_2') = P(E_1) [1 - P(E_2)]$$

$$\therefore U - A = A'$$

$$P(E_1 \cap E_2') = P(E_1) \cdot P(E_2')$$
 Proved.

Theorem 2:-

If E_1 & E_2 are Independent than show that E_1' & E_2 are also Independent:-

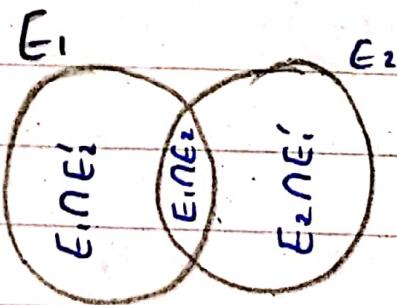
Proof:

Given:

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2)$$

To Prove:

$$P(E_1' \cap E_2) = P(E_1') \cdot P(E_2)$$



A/c to figure

$$P(E_1' \cap E_2) = P(E_2) - P(E_1 \cap E_2)$$

A/c to given Condition E_1 & E_2 are Independent

$$P(E'_1 \cap E_2) = P(E_2) - P(E_1)P(E_2)$$

$$P(E'_1 \cap E_2) = P(E_2)[1 - P(E_1)]$$

$$\because U - A = A'$$

$$P(E'_1 \cap E_2) = P(E_2) \cdot P(E'_1) \quad \text{Proved.}$$

Theorem 3:-

If E_1 & E_2 are Independent than
Show that E'_1 & E'_2 are also
Independent:-

Proof:

Given:

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2) \rightarrow ①$$

To prove:-

$$P(E'_1 \cap E'_2) = P(E'_1) \cdot P(E'_2) \rightarrow ②$$

By applying De Morgan's law on eq, ②
we have.

$$P(E'_1 \cap E'_2) = P(E_1 \cup E_2)'$$

$$\therefore U - A = A'$$

$$\rightarrow P(E_1 \cap E_2') = 1 - P(E_1 \cup E_2)$$

we know that

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

$$\rightarrow P(E_1 \cap E_2') = 1 - [P(E_1) + P(E_2) - P(E_1 \cap E_2)]$$

$$\rightarrow P(E_1 \cap E_2') = 1 - P(E_1) - P(E_2) + P(E_1 \cap E_2)$$

$$\rightarrow P(E_1' \cap E_2') = [1 - P(E_1) - P(E_2) + P(E_1 \cap E_2)]$$

from eq ①

$$\rightarrow P(E_1' \cap E_2') = [1 - P(E_1)] - P(E_2) + P(E_1) \cdot P(E_2)$$

$$\rightarrow P(E_1' \cap E_2') = [1 - P(E_1)] - P(E_2)[1 - P(E_1)]$$

$$\rightarrow P(E_1' \cap E_2') = [1 - P(E_1)] \cancel{+} [1 - P(E_2)]$$

$$\therefore U - A = A'$$

$$P(E_1' \cap E_2') = P(E_1') \cdot P(E_2')$$
 Proved.

PROBLEMS:-

A Firm has two operating systems working independently the probability that a specific system is available when needed is 0.96

a) what is the probability that neither

is available when needed.

b) what is the probability that an OS is available when needed.

Solution:-

CONDITION 1:-

$$P(O_1) = P(O_2) = 0.96$$

$$P(O_1') = P(O_2') = 1 - 0.96$$

$$P(O_1') = P(O_2') = 0.04$$

The probability of neither is available

$$P(O_1' \cap O_2') = P(O_1') \cdot P(O_2')$$

$$P(O_1' \cap O_2') = 0.04 * 0.04$$

$$P(O_1' \cap O_2') = (0.04)^2$$

Condition 2:-

$$P(O_1) = P(O_2) = 0.96$$

The probability of one of them is available.

$$P(O_1 \cup O_2) = P(O_1) + P(O_2) - P(O_1 \cap O_2)$$

$$P(O_1 \cup O_2) = P(O_1) + P(O_2) - P(O_1)P(O_2)$$

A/c to Independent condition

$$\frac{P(O_1 \cup O_2)}{P(O_1 \cup O_2)} = 0.96 + 0.96 - (0.96)^2$$

$$P(O_1 \cup O_2) = 0.9984$$

X ————— X ————— X

PROBABILITY:-

ADDITION RULE

MULTIPLICATION RULE

1. Mutually Exclusive 1. Independent

$$P(A \cup B) = P(A) + P(B) \quad P(A \cap B) = P(A) \cdot P(B)$$

2. Non Mutually Exclusive 2. Dependent

$$P(A \cap B) = P(A) \cdot P(B|A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

X ————— X ————— X

Conditional Probability:-

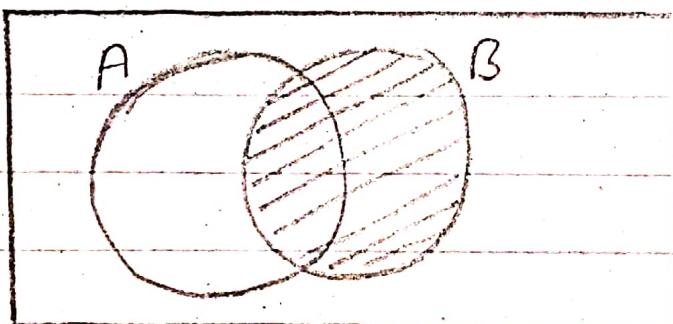
Conditional probability is defined as the likelihood of an event or outcome occurring based on the occurrence of a previous event or outcome.

$$\therefore P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$\therefore P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Let a Condition:

Find Probability of A given that
B is occurred



Hence it is a conditional probability
and Event B is occurred so the
sample space becomes

$$S = \{B\}$$

Now According to this sample space
we have to find probability of
Event A gives as required answer.

BAYES THEOREM:-

Bayes Theorem is nothing but the application of Conditional Probability.

According to Conditional Probability.

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

$$\boxed{P(A \cap B) = P(A|B) \cdot P(B)} \quad \textcircled{1}$$

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

$$\boxed{P(B \cap A) = P(B|A) \cdot P(A)} \quad \textcircled{2}$$

Comparing eq $\textcircled{1}$ & $\textcircled{2}$

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

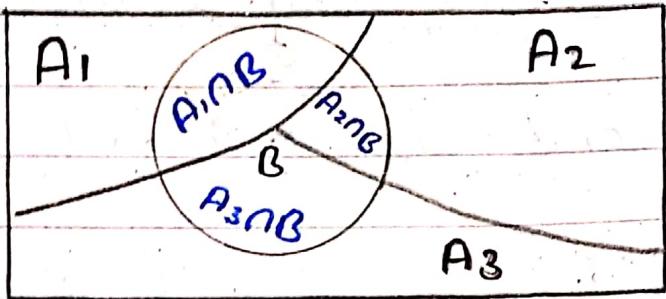
$$\boxed{P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}} \quad \textcircled{3}$$

↓

BAYES Theorem:-

Now we find a generalized case of Bayes Theorem.

Generalized Case Of Bayes' Theorem:-



$$A_1 + A_2 + A_3 = 100\%$$

$$P(A_1) + P(A_2) + P(A_3) = 1$$

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) \quad \textcircled{4}$$

we know that A/c to Conditional Probability

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

$$\therefore P(A \cap B) = P(A|B) \cdot P(B)$$

$\textcircled{4} \Rightarrow$

$$P(B) = P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + P(B|A_3) \cdot P(A_3)$$

$$P(B) = \sum_{i=1}^n P(B|A_i) \cdot P(A_i) \quad \textcircled{5}$$

from Eq \textcircled{3}

$$\rightarrow P(A_k|B) = \frac{P(B|A_k) \cdot P(A_k)}{P(B)}$$

A/c to eq ⑤

$$P(A_k|B) = \frac{P(B|A_k) \cdot P(A_k)}{\sum_{i=1}^n P(B|A_i) \cdot P(A_i)}$$

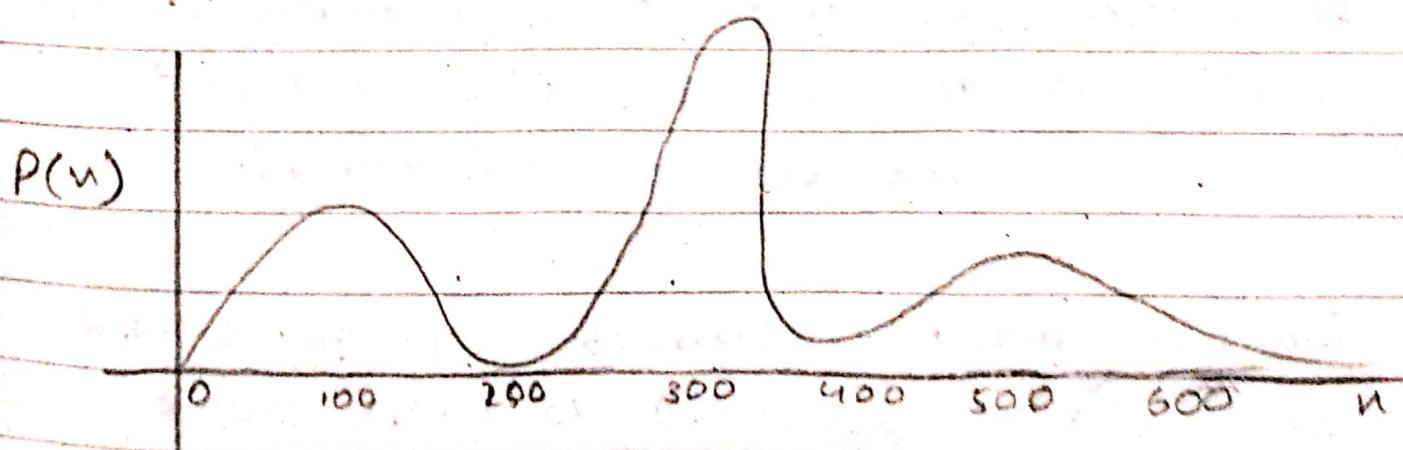
Generalized Case of Bayes theorem.

X ————— X ————— X ————— X

PROBABILITY DENSITY FUNCTION:-

Everything that can be known about random variable is contained in its Probability Density Function.

It allow us to determine the probability that a particular measurement of a random variable might be written some set of values.



Overview: Use Of Probability:

In any statistical analysis we use probability in two ways

1. Probability models describe non-deterministic nature of measurements
2. Probability is used to quantify the uncertainty in the results (conclusions) of our statistical analysis.

Random Variables:-

A Random Variable is a function of sample space, It is basically a device for transferring probability from complicated Sample spaces to simple Sample space.

A Random Variable X is a function whose domain is the sample space and whose range is the set of real numbers.

Thus a Random Variable assigns a real value (i.e. a number) to every outcome

in the sample space the particular value are called realisations and are denoted as x .

If the realisations are countable, n_1, n_2, \dots , the random variable is said to be discrete.

If there are many infinite uncountable realisation, the random variable

Random Variable can be classified according to the types of values they can take on the range of the random variable.

1. Binary
2. Categorical (Ordered Or Unordered)
3. Qualitative
4. Quantitative.

Probability Mass Function:-

If x is a discrete random variable the function given by $f(n) = P(x = x)$ for every n with in range of x , is

called probability mass function or probability function or probability distribution of X .

A function can serve the probability of a discrete random variable X iff its value $f(n)$ satisfy the condition

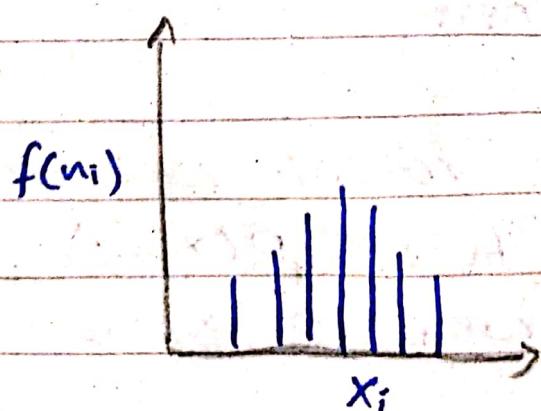
1. $f(n) \geq 0$
2. $\sum f(n) = 1$

Probability Distributions:- (Discrete R.V)

Table, formula or graphical representation of probability of random variable X

x_i	n_1	n_2	n_k	Total
$P(x_i)$	$P(n_1)$	$P(n_2)$	$P(n_k)$	1

Probability Histograms-



Probability Distribution:- (continuous R.V)

The probability distribution of a random variable describes how the probabilities are distributed. For a Continuous Random Variable outcomes and related probabilities are not defined at specific values, but rather over an Interval of values.

Example:

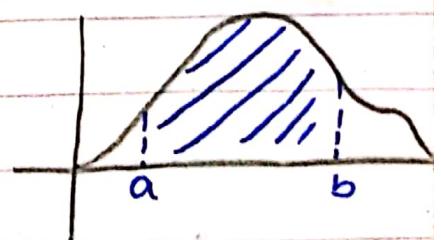
A random variable X , the weight of an adult blue crab caught from Ireland, may range of 0.5 to 0.3 kg the probability that an adult blue crab weighs between these values is the area under the curve of the probability density function.

Formulas:

$$P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{+\infty} f(u) du = 1$$

$$P\{X \in (a, b)\} = \int_a^b f(u) du =$$

$$P\{X < a\} = \int_{-\infty}^a f(u) du$$



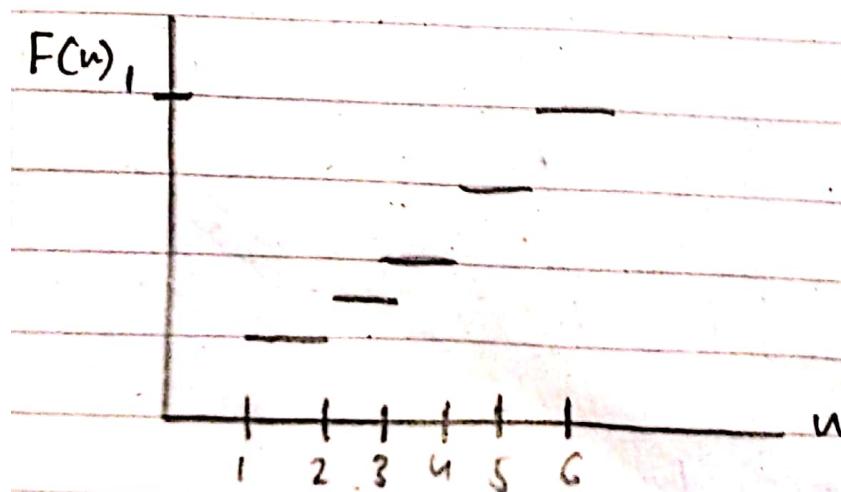
Commulative Distribution Function: (CDF)

The Commulative Distribution Function (cdf) is $F(n)$ of a discrete random variable X with p.m.f., $P(n)$ or $f(n)$ is defined for every number n by

$$F(n) = P(X \leq n) = \sum_{n=0}^n P(n)$$

For any number n , $F(n)$ is the probability that the observed value of X will be atmost n ,

For X a discrete Random Variable the graph will have jump at every possible value of X and flat between any two possible values of X such a graph is called Step Function:



For Continuous Random Variable.

$$F(a) = P\{u \leq a\} = \int_{-\infty}^a f(u) du$$

Example for C.d.f :-

Suppose a P.m.f of y is

y	1	2	3	4
$P(y)$	0.4	0.3	0.2	0.1

$$F(y) = \begin{cases} 0 & \text{if } y < 1 \\ 0.4 & \text{if } y \leq 2 \\ 0.7 & \text{if } y \leq 3 \\ 0.9 & \text{if } y \leq 4 \\ 1 & \text{if } y \leq 4 \end{cases}$$

X ————— X ————— X ————— X
Distribution of Boy's and Girl's having
three kids in a family.

$$n(\text{sample space}) = 2^3$$

$$n(S) = 8$$

case	1st	2nd	3rd	No of B
1	B	B	B	3
2	B	B	G	2
3	B	G	B	2
4	B	G	G	1
5	G	B	B	2
6	G	B	G	1
7	G	G	B	2
8	G	G	G	0

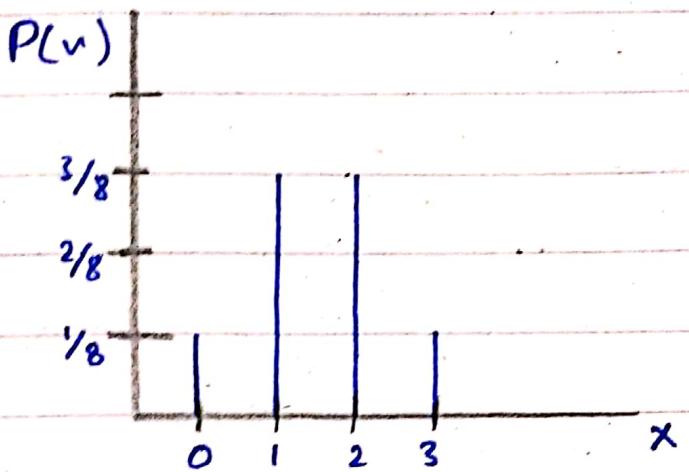
Tabular Form →

X	F	P(n)
0	1	1/8
1	3	3/8
2	3	3/8
3	1	1/8

Functional Form:-

$$P(n) = \frac{\binom{3}{n}}{2^3} ; n = 0, 1, 2, 3$$

Probability Histogram:-



Cumulative Density Function:-

$$P(X \leq n)$$

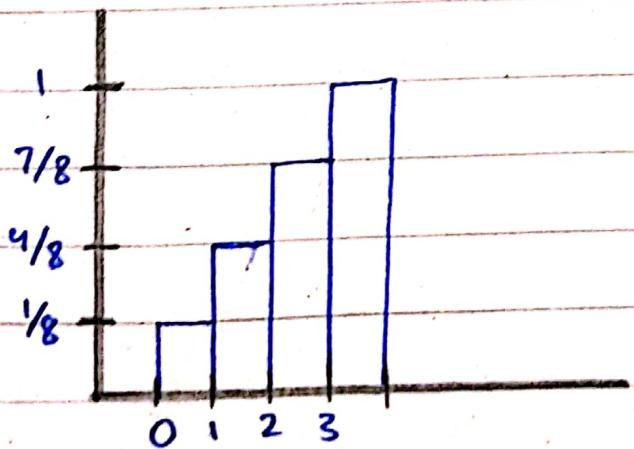
$$P(X \leq 0) = \frac{1}{8}$$

$$P(X \leq 1) = \frac{4}{8} = \frac{1}{2}$$

$$P(X \leq 2) = \frac{7}{8}$$

$$P(X \leq 3) = 1$$

STEP FUNCTION:-



Example For Continuous Random Variable:-

A Gas station operates two pumps, each of which can pump up-to 10,000 gallons of gas in a month. The total amount of gas pumped at the station in a month is a random variable y (measured in 10,000 gallons) with a p.d.f given by.

$$f(y) = \begin{cases} y & , 0 < y < 1 \\ 2-y & , 1 \leq y < 2 \\ 0 & , \text{else where} \end{cases}$$

a) Graph $f(y)$

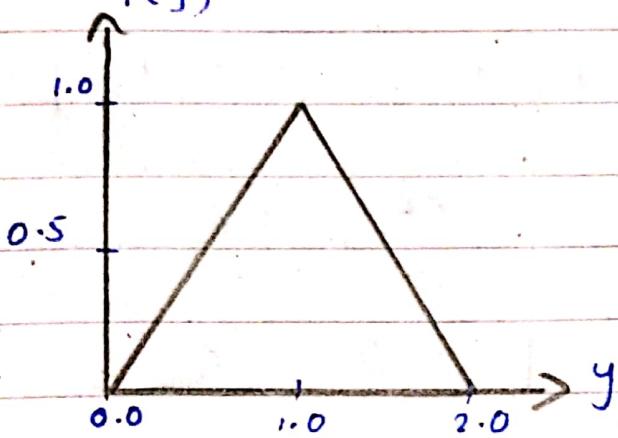
b) Find $F(y)$ and graph it

c) Find the probability that the station will pump b/w 8,000 and 12,000 gallons in a particular month.

d) Given that the station pumped more than 10,000 gallons in a month, find the probability the station pumped more than 15,000 gallons during the month.

Solution:-

a) $f(y)$



b)

$$F(y) = P(Y \leq y) = \int_{-\infty}^y f(y) dy$$

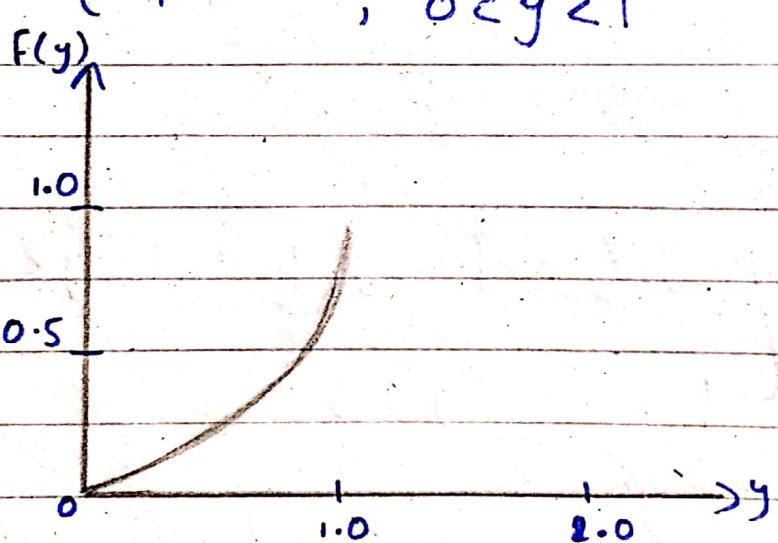
$$\Rightarrow F(y) = \begin{cases} \int_0^y y dy & ; 0 < y < 1 \\ \int_0^1 y dy + \int_1^y (2-y) dy & ; 1 \leq y < 2 \\ 1 & ; y \geq 2 \end{cases}$$

$$\Rightarrow F(y) = \begin{cases} \frac{y^2}{2} & ; 0 < y < 1 \\ \frac{y^2}{2} \Big|_0^1 + \int_1^y 2 dy - \int_1^y y dy & ; 1 \leq y < 2 \\ 1 & ; y \geq 2 \end{cases}$$

$$\Rightarrow F(y) = \begin{cases} \frac{y^2}{2} & ; 0 < y < 1 \\ y_2 + 2y \Big|_1^y - \frac{y^2}{2} \Big|_1^y & ; 1 \leq y < 2 \\ 1 & ; y \geq 2 \end{cases}$$

$$F(y) = \begin{cases} \frac{y^2}{2} & ; 0 \leq y < 1 \\ \frac{1}{2} + 2(y-1) - \frac{1}{2}(y^2-1) & ; 1 \leq y < 2 \\ 1 & ; y \geq 2 \end{cases}$$

$$F(y) = \begin{cases} \frac{y^2}{2} & ; 0 \leq y < 1 \\ 2y - \frac{y^2}{2} - 1 & ; 1 \leq y < 2 \\ 1 & ; y \geq 2 \end{cases}$$



$$c) P(8000 \leq y \leq 12,000) = P(0.8 \leq y \leq 1.2)$$

$$\begin{aligned} &= \int_{0.8}^1 y dy + \int_1^{1.2} (2-y) dy \\ &= \left[\frac{1}{2} \cdot y^2 \right]_{0.8}^1 + \left[2 \cdot y - \frac{1}{2} \cdot y^2 \right]_1^{1.2} \end{aligned}$$

$$= \frac{1}{2} \cdot (1 - (-0.8)^2) + 2(1 \cdot 2 - 1) - \frac{1}{2} ((1 \cdot 2)^2 - 1)$$

$$= 0.36$$

d) Using Conditional Probability.

$$\frac{P(Y > 15000 \cap X > 10000)}{P(X > 10,000)} = \frac{P(Y > 15000)}{P(X > 10,000)}$$

$$\frac{P(Y > 1.5)}{P(X > 1.0)} = \frac{1 - P(Y \leq 1.5)}{1 - P(X \leq 1.0)} = ①$$

For $P(Y \leq 1.5)$

$$P(Y \leq 1.5) = \int_0^1 y \, dy + \int_1^{1.5} (2-y) \, dy$$

$$P(Y \leq 1.5) = \left[\frac{1}{2} \cdot y \right]_0^1 + [2y]_1^{1.5} - \left[\frac{1}{2} y^2 \right]_0^{1.5}$$

$$P(Y \leq 1.5) = \frac{1}{2} \cdot (1 - 0) + 2(1.5 - 1) - \frac{1}{2} ((1.5)^2 - 1)$$

$$P(Y \leq 1.5) = 0.875$$

For $P(X \leq 1.0)$

$$P(X \leq 1.0) = \int_0^1 y \, dy = \left[\frac{1}{2} \cdot y^2 \right]_0^1$$

$$P(X \leq 1.0) = 0.5$$

$$\frac{P(Y > 15000)}{P(X > 10000)} = \frac{P(Y > 1.5)}{P(X > 1.0)} = \frac{1 - P(Y \leq 1.5)}{1 - P(X \leq 1.0)}$$

$$\frac{P(Y > 1.5)}{P(X > 1.0)} = \frac{1 - 0.875}{1 - 0.5} = 0.25 \text{ Ans}$$

MATHMATICAL EXPECTATION:-

Let 'X' be the random variable associated with probabilities $f(n)$ or $P(X=x)$.

The P.d.f is

X	x_1	x_2	x_k	Total
$P(x)$	$P(x_1)$	$P(x_2)$	$P(x_k)$	1

The mean or expected value of x is

when x is discrete

$$\mu = E(x) = \sum n P(n) =$$

when x is continuous

$$\mu = E(n) = \int_{-\infty}^{\infty} n P(n) dn$$

PROPERTIES:-

1. $E(x) = \mu = \bar{x}$

Proof

$$\begin{aligned} \text{As } \bar{x} &= \mu = \frac{\sum f_i x_i}{\sum f_i} = \frac{\sum x_i \cdot f_i}{N} \\ &= \sum x_i P(x_i) \\ \boxed{\bar{x}} &= E(x) \end{aligned}$$

2. $E(\text{constant}) = \text{constant}$

$$\begin{aligned} E(c) &= \sum c P(x_i) \\ &= c \sum P(x_i) \end{aligned}$$

$$\boxed{E(c) = c}$$

3. The variance of a random variable x

$$\text{is } \text{var}(x) = \sigma^2 = E(x^2) - \{E(x)\}^2$$

$$\text{var}(x) = E(x^2) - \mu^2$$

$$\sigma^2 = E\{(x-\mu)^2\} = E\{x^2 - 2\mu x + \mu^2\}$$

$$= E(x^2) - 2\mu E(x) + \mu^2$$

$$= E(x^2) - 2\mu^2 + \mu^2$$

$$\boxed{\sigma^2 = E(x^2) - \mu^2}$$

4. The expected value of the sum or difference of two or more function of random variable

X and Y is the sum or the difference
of the expected values of the functions

i.e

$$E(X \pm Y) = E(X) \pm E(Y)$$

Proof.

$$E[X \pm Y] = \sum_{\text{all } u} \sum_{\text{all } y} (u \pm y) \cdot P(u, y)$$

$$= \sum_x \sum_y (u) \cdot P(u, y) \pm \sum_u \sum_y y P(u, y)$$

$$= \sum_x u P(u) \pm \sum_y y P(y)$$

$$E[X \pm Y] = E(X) \pm E(Y)$$

5- If X and Y are discrete random variables
and a & b are constant then

$$E[ax \pm by] = aE(u) \pm bE(y)$$

Proof

Let

$$U = au, V = by$$

$$E[U \pm V] = ?$$

A/c to 4th property

$$E[U \pm V] = E(U) \pm E(V)$$

$$\rightarrow E[an + by] = E(an) + E(by)$$

$$\rightarrow E[an + by] = aE(n) + bE(y)$$

6. Let X & Y be the two Independent random Variable's then

$$E(XY) = E(X) \cdot E(Y)$$

Proof:-

$$E[XY] = \sum_x \sum_y xy \cdot P(x, y)$$

$\because X$ and Y are Independent

$$P(x, y) = P(x) \cdot P(y)$$

$$E[X \cdot Y] = \sum_x \sum_y xy \cdot P(x) \cdot P(y)$$

$$= \sum_x x P(x) \cdot \sum_y y P(y)$$

$$E[X \cdot Y] = E(X) \cdot E(Y)$$

$$7. E(g(u)) = \sum_{all x} g(u) \cdot P(u)$$

$$8. E(x^n) = \sum u^n P(u)$$

Raw Moments :- about origin

- When X is discrete R.V.

$$E(X^r) = \sum n^r P(x_i) = M_r$$

- When X is Continuous R.V.

$$E(X^r) = \int_{-\infty}^{\infty} n^r P(n) dn = M_r$$

PROBLEM:-

A Race car driver wishes to insure his car for the racing season's for \$50,000. The insurance company estimates a total loss may occur with probability 0.002, a 50% loss with probability 0.01, and a 25% loss with probability 0.1, Ignoring all other partial losses, what premium should the insurance company charge each season to realize an average profit of 500\$?

Sol

Let ' X ' be the profit to the company

and ' A ' be the amount of premium paid for every season.

Given $E(x) = \$500$

X	P(x)
A - 50,000 \$	0.002
A - 25,000 \$	0.01
A - 12,500 \$	0.1
A	0.888
Total	1

$$\therefore E(x) = \sum x P(x)$$

$$E(x) = (A - 50000)(0.002) + (A - 25000)(0.01) \\ + (A - 12500)(0.1) + A(0.888)$$

$$E(x) = 0.002A - 100 + 0.01A - 250 + \\ 0.1A - 1250 + 0.888A$$

$$E(x) = A - 1600 - ①$$

A/c to given data

$$\therefore E(x) = 500$$

① \Rightarrow

$$500 = A - 1600$$

$$\boxed{A = 2100 \$}$$

Decision:-

The company charge \$ 2100 every season in order to get an average profit of 500 \$

Problems:-

In a gambling game a man is paid 5\$ if he gets all heads or all tails when three coin's are tossed and he pays out \$3 if either one or two heads show. what is his expected gain?

Sol:

Let 'X' be the amount the gambler can win

when $X = 5\$$

$$E_1 = \{ HHH, TTT \}$$

when $X = -3\$$

$$E_2 = \{ HHT, HTH, THH, HTT, TTH, THT \}$$

$$P(E_1) = \frac{2}{8} = \frac{1}{4}$$

$$P(E_2) = \frac{6}{8} = \frac{3}{4}$$

Thus

X	5	-3	Total
$P(X)$	$\frac{1}{4}$	$\frac{3}{4}$	1

$$\text{Mean} = \mu = E(x) = \sum x P(n)$$

$$\sum x P(x) = 5(1/4) + (-3)(3/4)$$

$$\text{Mean} = \frac{5}{4} - \frac{9}{4}$$

$$\boxed{\text{Mean} = -1}$$

Decision:-

In this game gambler will en the average loss of 1\$ per toss of the three coins

Problem:

Suppose that the number of cars, x , that pass through a car wash b/w 4:00 pm and 5:00 pm on avg. Sunny Friday has the following Probability distribution.

X	4	5	6	7	8	9
$P(x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let $g(n) = 2n-1$ represent the amount of money in dollars, paid to the attendant by the manager. Find the attendant's expected earning's

For this particular time period.

Sol

By Property.

$$E[g(x)] = \sum g(u) f(u)$$

$$E[g(u)] = \sum (2u-1) f(x)$$

$$\begin{aligned} E[g(u)] &= (2(4)-1)(\gamma_{12}) + (2(5)-1)(\gamma_{12}) + \\ &\quad (2(6)-1)(\gamma_4) + (2(7)-1)(\gamma_4) + \\ &\quad (2(8)-1)(\gamma_6) + (2(9)-1)(\gamma_6) \end{aligned}$$

$$E[g(u)] = 12.67 \$$$

BERNOULLI DISTRIBUTION:-

Consider an experiment with two possible outcomes call them Success (s) and failure (f) [Alive or dead, Sweet or not sweet, Defective or Non-defective, Solved or Unsolved] with probabilities

$$P(s) = p \quad \& \quad P(f) = 1 - p = q$$

Let X be a discrete r.v take value 0 if failure occurs or 1 if success occur with probabilities.

$$P(X=0) = 1 - p = q$$

$$P(X=1) = p$$

The probability distribution of the values which the random variable X takes is given by.

$$P(X=n) = P^n (1-p)^{1-n} = p^n \cdot q^{1-n}; n=0,1$$

is called a Bernoulli distribution or Point Binomial distribution and the random var

is called a Bernoulli variable.

In Bernoulli Distribution's

no of Trials one
Success (P)
Failure ($1-P$) = q .

RAW MOMENTS:-

$$\therefore E(x^n) = P$$

Proof 1:

$$E(x^n) = \sum x^n P(x=n)$$

By the definition of Bernoulli distribution

$$E(x^n) = \sum x^n P^x (1-p)^{1-x}$$

$$E(x^n) = 1^n \cdot P^1 (1-p)^{1-1} + 0^n \cdot P^0 (1-p)^{1-0}$$

$$\boxed{E(x^n) = P}$$

Proof 2:

By Moment Generating Function of Bernoulli Distribution

$$\therefore M_n(t) = q + pe^t$$

$$E(x^r) = M_n^{(r)}(t)$$

$$M_n^{(r)}(t) = pe^t$$

Setting $t=0$

$$[E(x^r) = pe^0 = p]$$

Moment Generating function:-

$$P(x=n) = \begin{cases} q & : n=0 \\ p & : n=1 \\ 0 & : n \notin \{0,1\} \end{cases}$$

$$\therefore M_n(t) = E(e^{tx}) = \sum e^{tn} P(x=n)$$

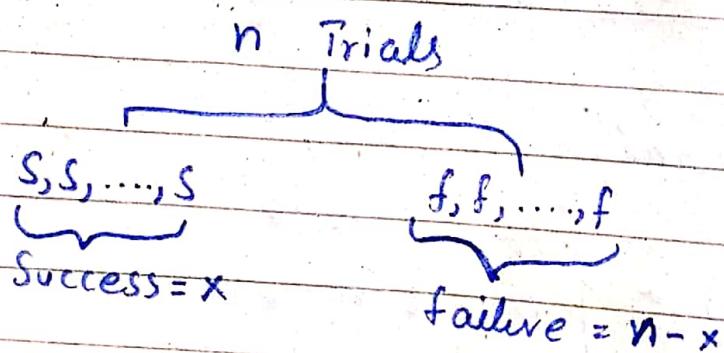
$$M_n(t) = e^{t(0)} P(x=0) + e^{t(1)} \cdot P(x=1)$$

$$M_n(t) = q + pe^t$$

Binomial Distributions:-

Consider an experiment with two possible outcome's call them success (S) and failure (f) with $P(S) = p$ & $P(f) = q$ such that $p+q=1$. Let 'X' be the random variable denote's the number of success in n independent repeat trials e.g.

Consider



The probability distribution of the particular sequence (by multiplicative law of Independent event's) is

$$p^x \cdot q^{(n-x)} \text{ or } p^x \cdot (1-p)^{(n-x)}$$

The number of sequence in which 'x' success and $n-x$ failures are observed in some order is $\binom{n}{x}$ ways. which is binomial coefficient

Thus the probability distribution that exactly X success and $n-x$ failures occur in n independent trials is

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}; x = 0, 1, 2, 3, \dots, n$$

which is known as binomial distribution with index ' n ' and parameter p

Q.

Prove That

$$\rightarrow \sum_{x=0}^n b(x; n, p) = 1$$

$$\rightarrow \sum_{x=0}^n \binom{n}{x} \cdot p^x (1-p)^{n-x}$$

$$\rightarrow \binom{n}{0} \cdot p^0 (1-p)^{n-0} + \binom{n}{1} \cdot p^1 (1-p)^{n-1} + \binom{n}{2} \cdot p^2 (1-p)^{n-2} + \dots + \binom{n}{n} \cdot p^n (1-p)^{n-n}$$

$$\rightarrow (1-p)^n + \binom{n}{1} p (1-p)^{n-1} + \binom{n}{2} \cdot p^2 (1-p)^{n-2} + \dots + p^n$$

$$\rightarrow q^n + \binom{n}{1} p \cdot q^{n-1} + \binom{n}{2} p^2 q^{n-2} + \dots + p^n$$

$$\rightarrow [q + p]^n$$

$$\because q + p = 1$$

$$\rightarrow 1^n = 1$$

proved.

RAW

Moment's In Binomial Distributions-

Let 'x' be a random variable with the binomial distribution $b(n; n, P)$.

The moments about origin are given by the relation

$$M_r = E(x^r)$$

at $r=1$

$$M_1 = E(x) = \sum_{x=0}^n x b(n; n, P)$$

$$= \sum_{x=0}^n x \cdot \binom{n}{x} \cdot p^x q^{n-x}$$

$$= (0) \cdot \binom{n}{0} \cdot p^0 q^{n-0} + (1) \cdot \binom{n}{1} \cdot q^1 \cdot p^{n-1} + (2) \cdot \binom{n}{2} \cdot p^2 q^{n-2} \\ \dots \dots + \cancel{(n) \cdot \binom{n}{n} \cdot p^n q^{n-n}}$$

$$= \sum_{x=0}^n x \cdot \frac{n(n-1)!}{(n-x)!(x-1)!} \cdot p^x q^{n-x}$$

$$= np \sum_{x=0}^n \frac{(n-1)!}{((n-1)-(x-1))! \cdot (x-1)!} \cdot p^{x-1} \cdot q^{n-x}$$

$$= np \sum_{x=0}^n \binom{n-1}{x-1} \cdot p^{x-1} \cdot q^{n-x}$$

$$= np [q^{n-1} + p]$$

$$\mu'_1 = np$$

at $r=2$

$$\mu'_2 = E[x^2] = \sum_{x=0}^n x^2 \cdot \binom{n}{x} \cdot p^x \cdot q^{n-x}$$

$$\mu'_2 = E[x^2] = E[x(x-1) + x]$$

$$\mu'_2 = E[x(x-1) + x]$$

$$\mu'_2 = E[x(x-1)] + E[x] - ①$$

$$E[x(x-1)] = \sum_{x=0}^n x(x-1) \cdot \binom{n}{x} p^x \cdot q^{n-x}$$

$$= \sum_{x=0}^n x(x-1) \cdot \frac{n(n-1)(n-2)!}{(n-x)(x-1)(x-2)!} \cdot p^x \cdot q^{n-x}$$

$$= n(n-1) \cdot p^2 \sum_{x=0}^n \frac{(n-2)!}{(n-x)(x-2)!} p^{x-2} \cdot q^{n-x}$$

$$= n(n-1)p^2 \sum_{x=0}^{n-2} \binom{n-2}{x-2} \cdot p^{x-2} \cdot q^{n-x}$$

$$= n(n-1)p^2 [q + p]$$

$$E[x(x-1)] = n(n-1)p^2$$

① =>

$$\mu'_2 = n(n-1)p^2 + np$$

$$\mu'_2 = n^2p^2 - np^2 + np$$

$$M_2' = np[nP - p + 1]$$

$$M_2' = np[np + 1 - p]$$

$$\therefore q = 1 - p$$

$$M_2' = np[np + q]$$

$$M_2' = npq + (np)^2$$

at $r = 3$

$$M_3' = E(x^3) = E[x(x-1)(x-2) + 3x(x-1) + x]$$

$$M_3' = E[x(x-1)(x-2)] + E[3x(x-1)] + E[x]$$

$$M_3' = E[x(x-1)(x-2)] + 3E[x(x-1)] + E[x]$$

A/c to above solution's:-

$$M_3' = E[x(x-1)(x-2)] + 3n(n-1)p^2 + np \quad \text{--- (1)}$$

For

$$E[x(x-1)(x-2)] = \sum_{x=0}^n x(x-1)(x-2) \cdot \frac{n!}{(n-x)!} \cdot p^{x-n} \cdot q^{n-x}$$

$$\Rightarrow \sum_{x=0}^n x(x-1)(x-2) \cdot \frac{n(n-1)(n-2)(n-3)!}{(n-x)!} \cdot p^3 \cdot p^{x-3} \cdot q^{n-x}$$

$$\Rightarrow n(n-1)(n-2) p^3 \cdot \sum \frac{(n-3)!}{(n-x)!(x-3)!} \cdot p^{x-3} \cdot q^{n-x}$$

$$E[x(x-1)(x-2)] = n(n-1)(n-2)p^3 \sum \binom{n-3}{x-3} p^{x-3} \cdot \frac{n-1}{n}$$

$$E[x(x-1)(x-2)] = n(n-1)(n-2)p^3$$

\Rightarrow

$$\mu'_3 = n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np$$

at $r=4$

$$\mu'_4 = E[x^4] = E[x(x-1)(x-2)(x-3) + 6x(x-1)(x-2) + 7x(x-1) + x]$$

$$\mu'_4 = E[x(x-1)(x-2)(x-3)] + 6E[x(x-1)(x-2)] + 7E[x(x-1)] + E[x]$$

A/c to above solutions:-

$$\mu'_4 = E[x(x-1)(x-2)(x-3)] + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np \quad \text{--- (1)}$$

For

A/c to above Sequence.

$$E[x(x-1)(x-2)(x-3)] = n(n-1)(n-2)(n-3)p^4$$

\Rightarrow

$$\mu'_4 = n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np$$

MOMENTS ABOUT MEAN IN BINOMIAL DISTRIBUTION:-

$$\therefore M_r = \frac{1}{n} \cdot \sum (x_i - \bar{x})^r$$

For $r=1$

$$M_1 = 0$$

For $r=2$

$$\therefore M_2 = M'_1 - (M'_1)^2$$

$$M_2 = npq + n^2p^2 - n^2p^2$$

$$M_2 = npq$$

For $r=3$

$$M_3 = M'_3 - 3M'_2 M'_1 + 2M'_1^3$$

$$M_3 = n(n-1)(n-2)p^3 + 3(n(n-1)p^2) + np - 3\{(npq + (np))^2\} \\ (np) + 2(np)^3$$

$$M_3 = n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np - 3\{(npq + n^2p^2)\}.(n) \\ + 2n^3p^3$$

$$M_3 = (n^2 - n)(n-2)p^3 + 3n^2p^2 - 3np^2 + np - 3\{n^2p^2q + n^3p^3\} \\ + 2n^3p^3$$

$$M_3 = n^3p^3 - 2n^2p^3 - n^2p^3 + 2np^3 + 3n^2p^2 - 3np^2 + np -$$

$$3n^2p^2q - 3n^3p^3 + 2n^3p^3$$

$$\mu_3 = -3n^2p^3 + 2np^3 + 3n^2p^2 - 3np^2 + np - 3n^2p^2(1-p)$$

$$\mu_3 = np - 3n^2p^3 + 2np^3 + 3n^2p^2 - 3np^2 - 3n^2p^2 + 3n^2p^3$$

$$\mu_3 = np + 2np^3 - 3np^2$$

$$\mu_3 = np[1 - 3p + 2p^2]$$

$$\mu_3 = np[1 - 3p + 2p^2]$$

$$\mu_3 = np[2p^2 - p - 2p + 1]$$

$$\mu_3 = np[2p^2 - 2p + 1 - p]$$

$$\mu_3 = np[2p(p-1) - (p-1)]$$

$$\mu_3 = np[(p-1)(2p-1)]$$

$$\mu_3 = -np(1-p)(2p-1)$$

$$\mu_3 = -npq(2p-1)$$

$$\mu_3 = npq(1-2p) = npq(1-p-p)$$

$$\mu_3 = npq(q-p)$$