Lecture # 1st

- **Statistics.**

  Definition, Data, variables, Types of variables, Data Distribution, Frequency Distribution, Exploratory data analysis tools, Measure of Central Tendency & Dispersion, Moments, Skewness and Kurtosis.

- **Probability.**

  Definition, Random variables, Probability Distribution, Types of Probability Distribution, Commutative Probability Distribution, Laws of Probability, Independent event, Conditional Probability, Baye's Theorem.
  Discrete Probability Distribution, Binomial. Poisson distributions.
  Continuous Probability distribution, Normal Distribution, Moment generating function.
  Chebyshare Inequality. Time Series. Regression. Correlation (Simple, Partial & multiple

Scanned with CamScanner

- Books :-

* Introductory Statistics (3rd ed.)
by Sheldon M. Ross.

* Introductory Statistics (9th ed.)
by Walpole, Myer's, R.H . . . .

Definitions :-

Statistics :-

is defined as study of collecting information, presenting, analysing and drawing conclusions with the aid of mathematics and computers.
The most essential activity is perhaps the development of methods based on probability theory. This area include estimation of population parameter(s) and test of hypothesis based on probability theory.

Probability :-

A phenomenon is called random if

Scanned with CamScanner

The exact outcome is uncertain. The mathematical study of such randomness is called Theory of Probability.
A probability model has two essential pieces of its description:
Ω, the sample space, the set of possible
A event is a collection of outcomes of subsets of the sample space.

Scanned with CamScanner

Statistics :-

- D. The probability assigns a numerical value to each event.

- Data :-

- Data
  - Primary
  - Secondary

- Observations
  - Qualitative
    - Discrete
  - Quantitative
    - Discrete
    - Continuous

Data arrange in :
Form (Ascending Order)
Frequency array
Frequency distribution

Frequency Distribution :-

Important characteristics of large mass of data can be readily assessed by grouping data into different classes and determining no. of observations that fall in each of classes. Such an arragement in tabular form is called frequency distribution.

Some important terms related to frequency distribution are:-

1. No of Classes (m) :-

* Sturge's Rule.

$\therefore m = 1 + 3.3 \log_{10} N \rightarrow$ arbitrary.

* Square root Rule.

$\therefore m = \sqrt{N}$.

2. Range :-

$\therefore$ Range = Max. value − Min. value.

3. Class Interval :-

$$\therefore C.I. = \frac{\text{Range}}{\text{No of Classes}} \Rightarrow \frac{R}{m}.$$

Scanned with CamScanner

4. The starting point :-

eg:-

Frequency Distribution for the Weights of 50 pcs of Luggage:

Weight (in Kg.):	No. of Pcs.
7 - 9	2
10 - 12	8
13 - 15	14
16 - 18	19
19 - 21	7

* Class Limits :-

The smallest and largest values that can fall in a given class interval are referred to as its **class limits.**

eg. in interval 10 - 12.

10 → smaller no. : Lower class limit.

12 → larger no. : Upper class limit.

Scanned with CamScanner

* Class Boundry :-

* Class Frequency :-

No. of observations falling in a particular class is called class frequency (f).

* Class width :-

Numerical difference between the upper and lower class boundaries of a class interval is called class width (c).

eg. $c = 3$ in above data.

* Class mark or midpoint :-

The midpoint between upper and lower class boundaries or limits is called class mark or midpoint.

eg.

Scanned with CamScanner

Class Interval:	Class Boundries:	Class (x) Mark/Mid.	Frequency: f
7 - 9	6.5 - 9.5	8	2
10 - 12	9.5 - 12.5	11	8
13 - 15	12.5 - 15.5	14	14
16 - 18	15.5 - 18.5	17	19
19 - 21	18.5 - 21.5	20	7

Scanned with CamScanner

20-Jan-2020

Lecture # 2 :

Problem :-

Consider the data of lives of 40 similiar car batteries.

2.2	4.1	3.5	4.5	3.2	3.7	3.0	2.6
3.4	1.6	3.1	3.3	3.8	3.1	4.7	3.7
2.5	4.3	3.4	3.6	2.9	3.3	3.9	3.1
3.3	3.1	3.7	4.4	3.2	4.1	1.9	3.4
4.7	3.8	3.2	2.6	3.9	3.0	4.2	3.5.

Construct frequency distribution and comment on shape of distribution graphically :

Solution :

→ no. of classes:

$\therefore m = 1 + 3.3 \log_{10} N.$ $\qquad$ $n = 40$ (Given)

$m = 6.286$

$m \approx 7.$

→ Range.

$\therefore R = Max. - Min.$ $\qquad$ $max = 4.7$

$R = 4.7 - 1.6.$ $\qquad$ $min = 1.6$

Scanned with CamScanner

$R = 3.1$

→ Class Interval:

$\therefore C.I = \frac{R}{m}$

$C.I = \frac{3.1}{7}$

$c / C.I = 0.442 \Rightarrow 0.5 = c$

→ The starting point:

Lowest C.I is = 1.5.

so, Lowest C.B will be = 1.45 ; Lower C.L = 1.5

$\therefore c = 0.5$; Upper C.L = 1.5 + 0.4 =

Highest C.B will be = 1.45 + 0.5 = 1.95.

→ midpoints:

$\therefore$ (Upper C.L + Lower C.L) ÷ 2.

$\Rightarrow \frac{1.5 + 1.9}{2} = 1.7$

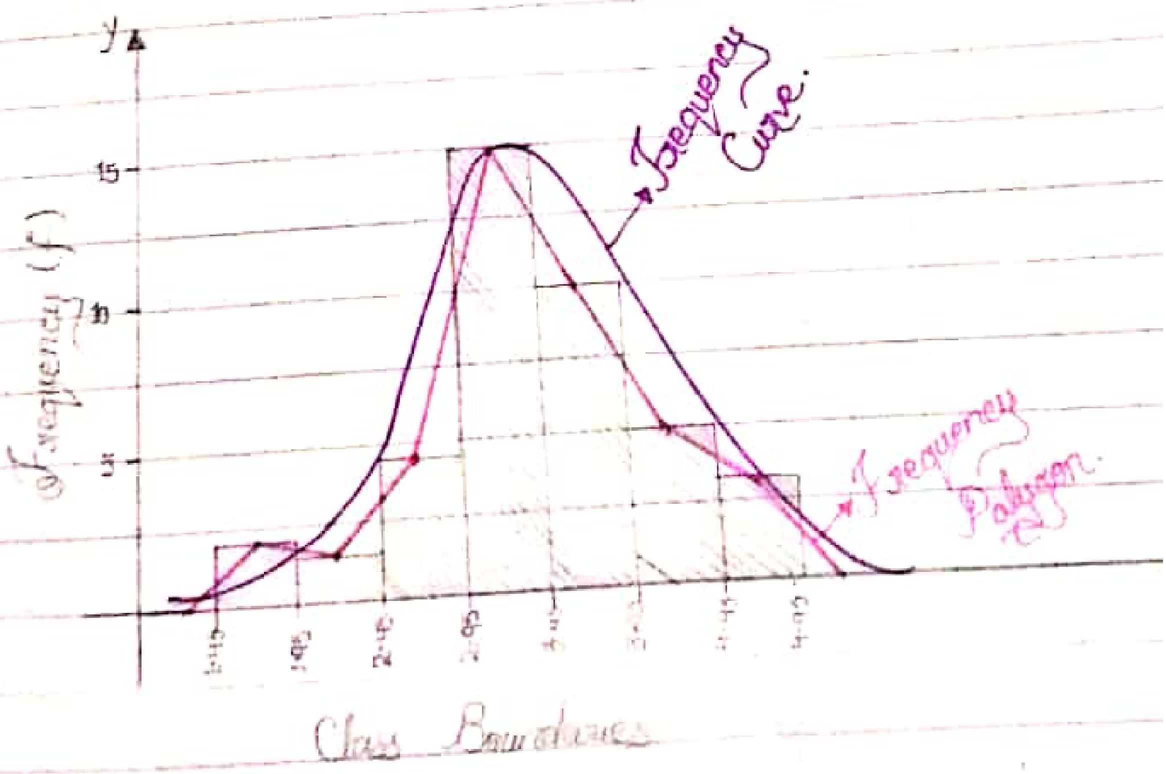
* C.F: Commutative Frequency.
* R.F: Relativ Frequencies or percentages.

$\therefore \frac{f}{T.f}$ or $\therefore \frac{f}{T.f} \times 100\%. \Rightarrow R.F \times 100.$

$\Rightarrow \frac{2}{40} = 0.05$ $\Rightarrow 5\%.$

Class Limit.	Class Boundary.	Mid Point.	Tally Marks.	Freq. f.	C.F.	R.F.
1.5 – 1.9	1.45 – 1.95	1.7	II	2	2	0.05
2.0 – 2.4	1.95 – 2.45	2.2	I	1	2+1=3	0.075
2.5 – 2.9	2.45 – 2.95	2.7	IIII	4	3+4=7	0.175
3.0 – 3.4	2.95 – 3.45	3.2	IIIIIIIIIIIIIII	15	7+15=22	0.55
3.5 – 3.9	3.45 – 3.95	3.7	IIIIIIIIII	10	22+10=32	0.8
4.0 – 4.4	3.95 – 4.45	4.2	IIIII	5	32+5=37	0.925
4.5 – 4.9	4.45 – 4.95	4.7	III	3	37+3=40	1
				40		

Graph :-



* Assumptions:
  - Class frequency concentrated at midpoint.
  - Class Interval evenly distributed among whole class.

Scanned with CamScanner

Measure of Central Location/Tendency :-

* average.

An average is the measure of center of a set of data when data are arranged in an increasing or decreasing order of magnitude. (a single value representing whole data)

Any measure indicating the centre of the set of data, arranged in an increasing or decreasing order of magnitude is called a measure of central location/tendency.

and,
The most commonly used measures of central location are :

- mean
- median
- mode

* Types of Means :-

- Arithmetic Mean, Median, Mode } mostly used.
- Geometric mean } for %ages
- Harmonic mean } for %ages
- Weighted mean
- Trim mean

Scanned with CamScanner

Arithmetic Mean (AM) :-

For Data :-

Ungrouped :-

Grouped :-

* Mean :-

Ungrouped:

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

$n \to$ no. of observations
$x_i \to$ data values
$\sum \to$ summation (sum)

Grouped:

$$\left(\sum_{i=1}^{k} f_i x_i\right) \div \left(\sum_{i=1}^{k} f_i\right)$$

$x_i \to$ mid-points
$f_i \to$ frequency

* Mode :-

Ungrouped:

∴ most repeated value
data may be :
uni-moded → 1 mode
bi-moded → 2 modes
tri-moded → 3 modes
no modes

Grouped:

$$\therefore X_L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h$$

$X_L \to$ median class
$f_m \to$ max. frequency
$f_1 \to$ frequency above median class.
$f_2 \to$ frequency below median class.
$h \to$ class interval.

Scanned with CamScanner

Median :-

if n is:

odd :

$\left(\frac{n+1}{2}\right)^{th}$ value of ordered data.

even :

$\left(\frac{n}{2}\right)^{th} + \left(\frac{n+2}{2}\right)^{th}$ value of ordered data.

$\therefore X_L + \frac{(n/2 - C)}{f} \times h.$

$X_L \rightarrow$ median class's upper class limit.

$C \rightarrow$ no. of classes.

$f \rightarrow$ max frequency.

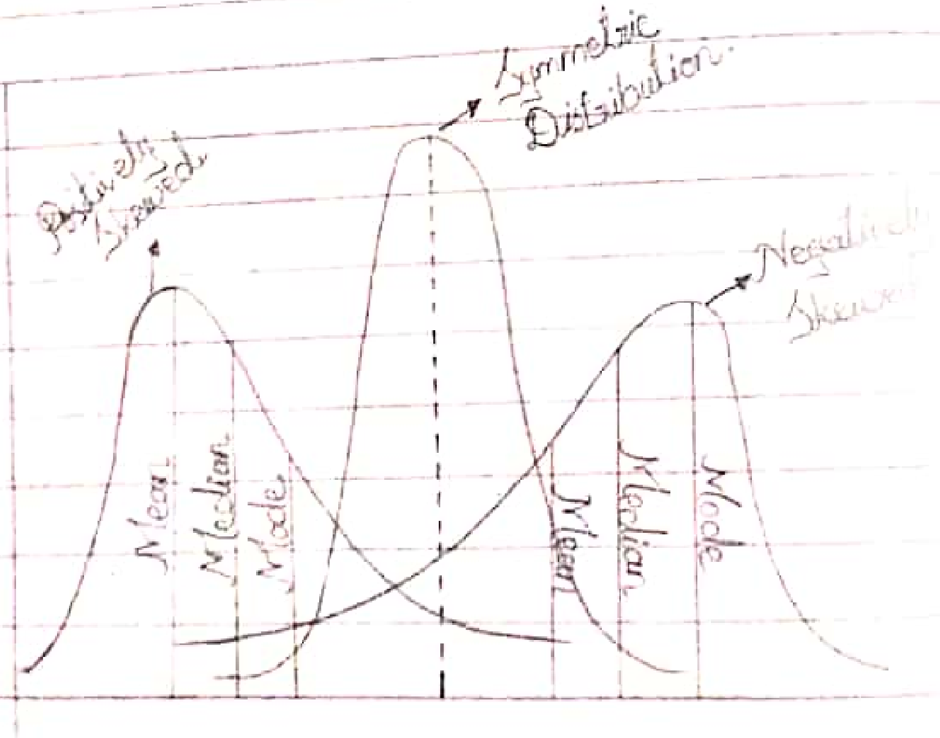
$h \rightarrow$ C.I.

Scanned with CamScanner

Frequency Curves nature by Shape :-

Data Shapes can be of three types :

1. Symmetric Distribution.
2. Positively Skewed.
3. Negatively Skewed.



Frequency Curves.

1. Symmetric Distribution :-

In S.D.

Scanned with CamScanner

∴ mean = mode = median.

eg. height of children.

2. Negatively Skewed :-

In N.S (Long-tail towards left).

∴ mode > median > mean.

eg. result if exam is easy.

3. Positively Skewed :-

In P.S (Long-tail towards right).

∴ mode < median < mean.

eg. result if exam is difficult.

Scanned with CamScanner

* **Measure of Dispersion** i.e

Dispersion is a word that measure the variability in values of unit and is define as degree of scatterness or clusterness of observation of given data from its central value.

* **Fractiles or Quantiles** :-

Measures of location that describe or locate the position of certain non-central pieces of data relative to entire set of data; are referred to as fractiles or quantiles.

1. Percentiles.
2. Deciles.
3. Quartiles.

* **Quartiles** :-

Quartiles are values that divide data (a set of observations) into four equal parts.

$\therefore Q_i = \left(\frac{in}{4}\right)^{th}$ value of ordered data.

where:

$i = 1, 2, 3, 4.$

Scanned with CamScanner

$\Rightarrow Q_1 = \frac{n}{4}$

$\Rightarrow Q_2 = \frac{2n}{4} \Rightarrow \frac{n}{2}$.

$\Rightarrow Q_3 = \frac{3n}{4} \Rightarrow \frac{3n}{4}$.

$\Rightarrow Q_4 = \frac{4n}{4} \Rightarrow n$.

* Deciles :-

Deciles are values that divide data (a set of observations) into ten equal parts.

$$\therefore D_i = \left(\frac{in}{10}\right)^{th} \text{ value of ordered data.}$$

where :

$i = 1 - 10$.

* Percentiles :-

Percentiles are values that divide data (a set of observations) into hundred equal parts.

$$\therefore P_i = \left(\frac{in}{100}\right)^{th} \text{ value of ordered data.}$$

Scanned with CamScanner

Deviations or Variance :-

Types of deviations are:
1. Range Deviation.
2. Quartile Deviation.
3. Mean Deviation.
4. Standard Deviation.

if
⇒

• f

4.

1. Range Deviation :-

$$\therefore R.D = \frac{x_m - x_0}{2}.$$

2. Quartile Deviation :-

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

where: $Q_3 = \frac{3n}{4}$ &
$Q_1 = \frac{n}{4}$.

3. Mean Deviation :-

• For Ungroup data :-

$$\therefore M.D_c = \frac{1}{n} \sum |x_i - c|.$$

where c = mean or median or mode.
[can put any of above].

Scanned with CamScanner

if $c = \text{mean}(\bar{x})$.

$\Rightarrow M.D_{\bar{x}} = \frac{1}{n} \sum |x_i - \bar{x}|$.

- for Group data:-

$\therefore M.D_c = \frac{1}{n} \sum_{i=1}^{k} f_i |x_i - c|$.

4. Standard Deviation:

positive square root of average of the square deviation taken from mean.

$$\therefore S.D = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

- Sample Space Variance / Deviation:

$$\therefore S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} \Rightarrow S^2 = \frac{1}{n-1} \left[\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right]$$

- Population Standard Variance / Deviation:

$$\therefore \sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2} \quad ; \mu_c = \frac{\sum_{i=1}^{k} N_i \mu_i}{\sum_{i=1}^{k} N_i}$$

$$\Rightarrow \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2,$$

; $\mu_c \to$ Combined mean of population

Scanned with CamScanner

27-1-20.

L # 3 :-

* Moments:- (3)

Moments of a distribution is designated by the power of deviation and it varies toward at end.

Mathematically.

1 → * r^{th} moment about mean:-

$$\mu_r = \frac{1}{n} \Sigma (x_i - \bar{x})^r \quad \text{OR}$$

$$\mu_r = \frac{1}{n} \Sigma f_i (x_i - \bar{x})^r \quad \text{(Grouped data)}$$

→ when $r = 1$.

$$\mu_1 = \frac{1}{n} \Sigma (x_i - \bar{x})$$

deviation about mean is always least. i.e $= 0$.

$$= \frac{1}{n} \Sigma x_i - \frac{1}{n} \Sigma \bar{x} \qquad \Sigma \bar{x} = n\bar{x}$$

$$= \frac{1}{n} \Sigma x_i - \frac{1}{n} n\bar{x}$$

$=$

$= \bar{x} - \bar{x} = 0$

Scanned with CamScanner

when $r=2$:

$\mu_2 = \mu_2' - \mu_1'^2$

→ Moments about origin or
Raw Moments :-

→ r^{th} moment about origin:-

$\mu_r' = \frac{1}{n} \sum x_i^r$

$r=1, \Rightarrow \mu_1' = \frac{1}{n} \sum x_i = \bar{x}$

$r=2, \Rightarrow \mu_2' = \frac{1}{n} \sum x_i^2$

$r=3, \Rightarrow \mu_3 = \frac{1}{n} \sum x_i^3$

$r=n \Rightarrow \mu_n' = \frac{1}{n} \sum x_i^{n/4}$

→ when $r=3$.

$\mu_3 = \mu_3' - 3\mu_2'\mu_1'^2 + 2\mu_1'^3$.

→ when $r=4$.

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$.

Scanned with CamScanner

3. ☆ For Moment about Constant:-

$\bar{x} = c$

→ r^{th} moment about constant 'c':-

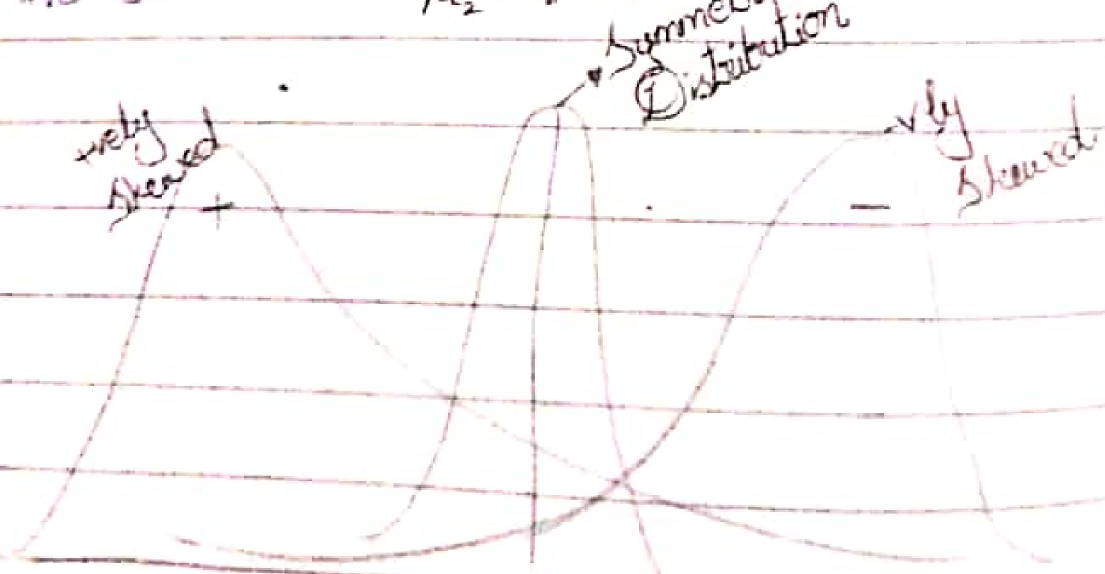
$\mu_r = \frac{1}{n} \sum (x_i - c)^r$

☆ Skewness:-

The degree of departure of frequency curve of a data distribution from symmetric distribution is called as Skewness. The statistics used for measuring the skewness is:

sq. quantity (+ve)

[is always +ve] ← $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$. variance (+ve)



$\mu_3 = \frac{1}{n} \Sigma (x_i - \bar{x})^3$

so if μ_3 is -ve ;
Then it is -vely skewed.

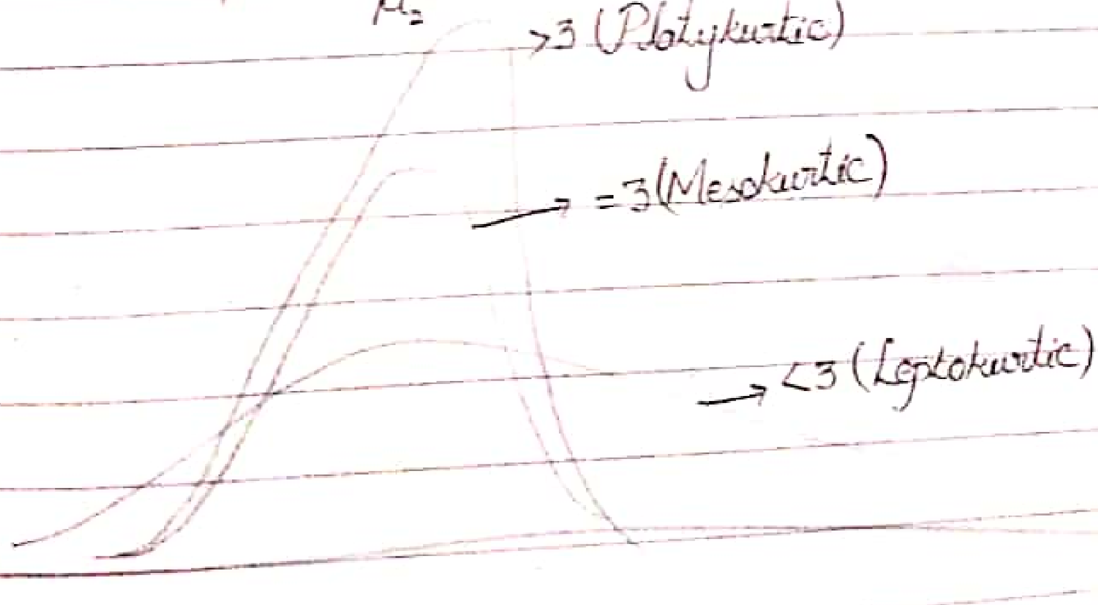
if β_1 = +ve (+vely skewed)
β_1 = -ve (-vely skewed)
β_1 = 0 (Symmetric)

☆ Kurtosis :-

It measures the peakedness & flatedness of data distribution.

Mathematically

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$



Scanned with CamScanner

→ β_2 :-

$\beta_2 > 3$; Platokurtic

$\beta_2 < 3$; Leptokurtic.

$\beta_2 = 3$; Mesokurtic

→ if :-

- $\beta_1 = 0$ & $\beta_2 = 3$ so ;

Normal or Gaussian Distribution

Scanned with CamScanner

* Coefficient of Variation (CV) :-

$CV = \frac{\sigma}{\mu} \times 100$

$CV = \frac{s}{\bar{x}} \times 100$

* CV < 50 ... less ...

* Exploring Data Analysis Techniques :-
(Tukey 1977)

1- Stem and Leaf Plot
2- Box and Whiskers Plot.

Stem & Leaf Plot Method

e.g. Data (40 Observations)

2.2	4.1	3.5	4.5	3.2	3.7	3.0	2.6
2.4	1.6	3.1	3.3	3.8	3.1	4.7	3.7
2.5	4.3	3.4	3.6	2.9	3.3	3.9	3.1
3.3	3.1	3.7	4.4	3.2	4.1	1.9	3.4
4.7	3.8	3.2	2.6	3.9	3.0	4.2	3.5

2 - digit data

so, left digit — stem

right digit — leaf.

Calculations:-

Stem.	Leaf. [unordered plot]
1	6 9
2	2 4 5 6 9 6
3	3 1 8 5 1 4 7 2 3 6 2 8 2 9 7 1 3 0
4	7 1 3 5 4 1 7 2

Stem	Leaf:- [Ordered]
1	6 9
2	2 4 5 6 6 9
Median Class ✓ 3	0 $\underset{Q_1}{\underline{0}}$ 1 1 1 1 2 2 2 3 3 $\boxed{3\ \ 4}$ 4 5 5 6 7
4	1 1 2 3 4 5 7 7

$Q_2 = \frac{3.3 + 3.4}{2}$

$\frac{40}{2} = 20$ n...

gives idea of Histogram i.e Symmetric.

[Box & Whiskers Plot Method]..

2 Method:-

Quantiles: $\frac{n}{4} = \frac{40}{4} = 10$

$Q_1 = \frac{n}{4} = 3.0$

$Q_3 = \frac{3n}{4} = 30^{th} = 3.9$

$Q_2 = \frac{2n}{4} = 20^{th} = 3.3$

• IQR (Inter Quantile Range) = $Q_1 - 1.5(IQR)$

1 7 3 4 5 ∴ IQR = $Q_3 - Q_1$

•

17 8 8 9 9

Q_3

Lower Outer Fence

Lower Inner Fence

Upper Inner Fence

Upper Outer Fence

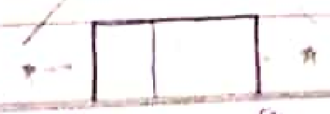
Suspected Outlier

1.6

Suspected Outlier

Highly Suspected Outlier

Whiskers



0(IQR) ... (0.9)

$Q_1 - 1.5(IQR)$
$= 3.0 - 1.5(0.9)$
$= 6.9 - 1.65$

Q_1 3.0 Q_2 3.35 Q_3 3.9

$Q_3 + 1.5(IQR)$
$= 3.9 + 1.5(0.9)$
$= 5.25$

$Q_3 + 3.0(IQR)$
$= 3.9 + 3.0(0.9)$
$= 6.65$

-vely Skewed

$Q_3 - Q_1 = IQR$.

L#4:-

3-Feb-20

Probability:

Definition:-

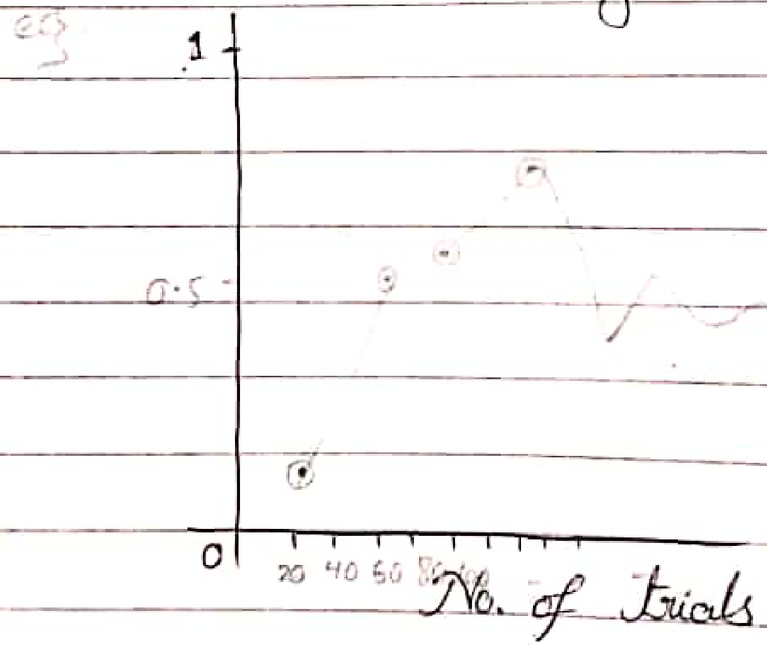
* is defined as the study of randomness
* depends on chances.
* is measure of beleif.

Cumulative
Comparative

$$\lim_{n\to\infty} P_r(E) = \frac{\text{Com no. of Heads}}{\text{Total no. of Trials}} \simeq 0.5$$

Graph:-

Commulative Probability.



* Probability lies b/w 0 & 1
  i.e we have $\frac{1}{2}$ chance.

Scanned with CamScanner

• for Probability :-
we should know.

• Sets:
a well defined collection of objects or items under consideration.

• Subset : → (E) U → Universal set.
a part of Universal (larger) set
$E \subset U$.

• Superset.
$U \supset E$.
larger set in which all subsets are defined (lies)

• Event :-
A sample point or a group of sample point of some particular interests is called and events.

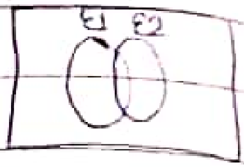
• Mutually Exclusive-Events :-
Two events are said to be mutually exclusive if the occurence of one precludes the occurence of other

Scanned with CamScanner

اگر ایک وقت میں ایک سے [illegible] occure [illegible] Irregular of [illegible]

if the simultanous occurence is not possible

2. Non-Mutually Exclusive Events :-



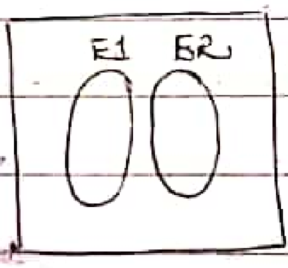
3. Independent Events :-

means dependent on each other.

eg.

3:- eg.

Venn Diagram



→ Disjoint events (eg: heads or tails (only one can occur at a time))

• Venn Diagrams :-

is a diagram related to set theory in mathematics by which the occurence of events can be visualised or portrayed.

Theorem :-

(TH1):- If E_1 & E_2 be the two events define on the sample space S, then the probability that E_1 occur or E_2 occur or atleast one of them occur is given by:

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

Proof :-

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

E_1 & $E_1' \cap E_2$ are mutually Exclusive

because

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

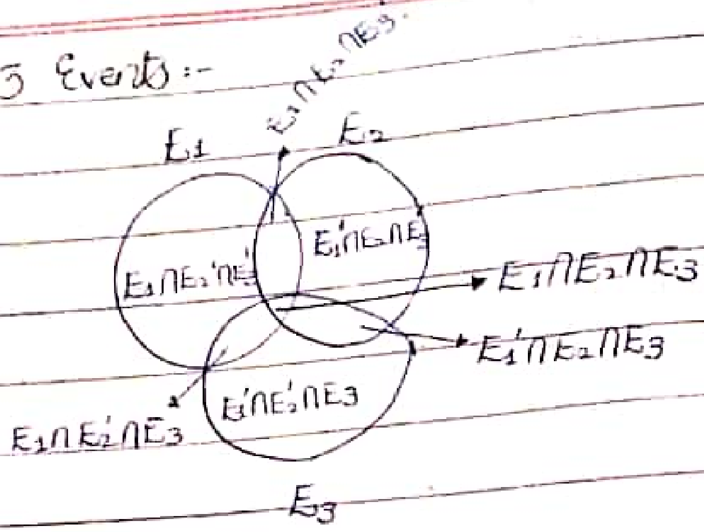
$$= P(E_1) + \underbrace{P(E_1' \cap E_2) + P(E_1 \cap E_2)}_{P(E_2)} - P(E_1 \cap E_2)$$

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

proved !

Scanned with CamScanner

- For 3 Events :-



Q#2 :-

So; same theorem for 3 events.-

If E_1, E_2 and E_3 be three events define on the sample space S, then the probability that E_1 occur or E_2 occur or E_3 occur or at least one of them occur is given by :-

$$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3) \quad \text{--- } 1$$

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

$$= P(E_1) + \underbrace{P(E_1' \cap E_2) + P(E_1 \cap E_2)}_{P(E_2)} - P(E_1 \cap E_2)$$

$$= P(E_1) + P(E_2) - P(E_1 \cap E_2) \quad \text{--- } (2)$$

let

$E = E_1 \cup E_2$ [from eqn (2)]

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

so;

$\Rightarrow P(E \cup E_3) = P(E) + P(E_3) - P(E \cap E_3)$ —— (3)

Substituting value from eqn (2) in (3)..

$P(E \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E \cap E_3)$ —— (4)

for $E \cap E_3$:-

$E \cap E_3 = (E_1 \cup E_2) \cap E_3$

$= E_1 \cap E_3 \cup E_2 \cap E_3$ [By Distributive law]

so:

$P(E \cap E_3) = P(E_1 \cap E_3 \cup E_2 \cap E_3)$.

$= P(E_1 \cap E_3) + P(E_2 \cap E_3) - P(E_1 \cap E_2 \cap E_3)$ —— (5)

Substitute (5) in (4) we have :-

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

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

Scanned with CamScanner

* If N events :-

Case 3

If $E_1, E_2, E_3, \ldots\ldots, E_N$ be N events

then :-

$$S_1 = \sum_{i=1} P(E_i)$$

$$S_2 = \sum_{i<j} P(E_i \cap E_j)$$

$$S_3 = \sum_{i<j<k} P(E_i \cap E_j \cap E_k).$$

$\vdots$

$$S_N = \sum_{i<j<k\cdots<N} P(E_1 \cap E_2 \cap \cdots \cap E_N).$$

So,

$P(E_1 \cup E_2 \cup \cdots \cup E_N)$

$$= S_1 - S_2 + S_3 - \cdots + (-1)^{N+1} S_N.$$

- **Independent Events :-**

Two events are said to be independent if the occurence of one doesn't effect on the occurence of other.

i.e.

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

OR if the intersection of both events is the product of events.

- **Theorem :-** OR E_1' and E_2. OR E_1' & E_2'

Q) If E_1 & E_2 are independents then show that E_1 and E_2' are also independent.

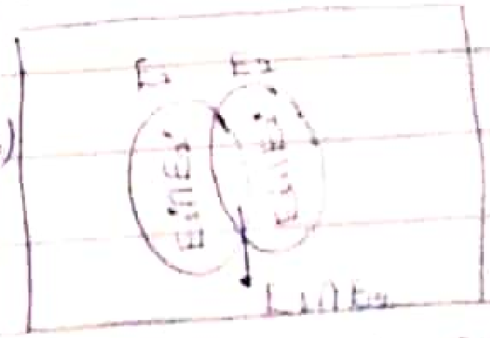
Proof :-

Given :-

$$P(E_1 \cap E_2) = P(E_1).P(E_2). \quad \text{——— } 1$$

To prove that :-

$$P(E_1 \cap E_2') = P(E_1).P(E_2'). \quad \text{——— } 2.$$



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

$$= P(E_1) - P(E_1).P(E_2)$$

$$= P(E_1)[1 - P(E_2)]$$

$$= P(E_1).P(E_2').$$

$\because P(E') = [1 - P(E)].$

Proved !

Q#2:- If E1 & E2 are independent then show that E1' & E2' are also independent.

Proof:-

Given:-

$P(E_1 \cap E_2) = P(E_1).P(E_2)$ ___ 1

to prove:

$P(E_1' \cap E_2') = P(E_1')P(E_2')$ ___ 2

By Demorgan's Law:-

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

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

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

$= 1 - P(E_1) - P(E_2) + P(E_1)P(E_2)$

$= 1[1 - P(E_1)] - P(E_2)[1 - P(E_1)]$

$= [1 - P(E_1)][1 - P(E_2)]$

$= P(E_1').P(E_2')$

Proved!

* Problem :-

Comp. Sci. department has two O.Ss working independently. The probability that a specific system is available when needed is 0.9.

a) What is Probability that neither is available when needed?

b) a system is available when needed?

a) $P(C_1) = 0.9$; $P(C_2) = 0.9$.

Soln:-

$P(C_1') = 1 - P(C_1)$

$= 1 - 0.9$

$= 0.1$.

$$P(C_1' \cap C_2') = P(C_1').P(C_2').$$

$$= (0.1)(0.1)$$

$$\Rightarrow P(C_1' \cap C_2') = 0.01 \quad \text{Answer}$$

b) - b/c C.Ss are independent.

$$P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2)$$

$$= (0.9) + (0.9) - P(C_1).P(C_2)$$

$$= (1.8) - (0.9)(0.9)$$

$$= 1.8 - 0.81.$$

$$= 0.99 \quad \text{Answer}$$

Scanned with CamScanner

L#5:-

10-Feb-20

Probability.

★Conditional Probability:-

The probability of any event depending on the outcome of another. The sample space of experiment might often be reduced when some additional information partaing to the outcome of experiment is received. Probability associated with such a reduced sample space is called Conditional Probability.

Mathematically,

If E_1 and E_2 are define on the sample space S & the occurence of the event E_2 depend on the occurence of event E_1 then,

$$P(E_2/E_1) = \frac{P(E_1E_2)}{P(E_1)} = \frac{P(E_1 \cap E_2)}{P(E_1)},$$

read as probability E_2 given E_1.

provided $P(E_1) > 0$

Scanned with CamScanner

Problem :-

a die shows 3 given that the die shows an odd number?

Solution :-

If a die is thrown

odd.

$S = \{1, 2, 3, 4, 5, 6\}$

$n(S) = 6$ case.

3 cases are odd.

E_1 : shows 3 , $n(E_1) = 1$.

E_2 : show odd no.

$n(E_2) = 3$.

[just 1 case to occur 3 q ki 3 ek hi bar hoy to ek hi time aa skta hay.]

$$\therefore P(E_1/E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

$$\text{"} = \frac{n(E_1 \cap E_2)/\cancel{n(S)}}{n(E_2)/\cancel{n(S)}}$$

$$\text{"} = \frac{n(E_1 \cap E_2)}{n(E_2)}$$

$\rightarrow n(E_1 \cap E_2) = 1$ [coz 1 common case in both.]

$\rightarrow n(E_2) = 3$

$$\Rightarrow \boxed{P(E_1/E_2) = \frac{1}{3}}$$

Answer!

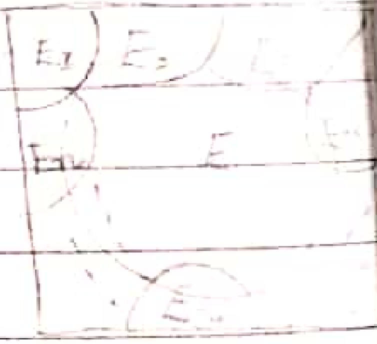
Scanned with CamScanner

Baye's Theorem :-
[Generalization of C.P].

If $E_1, E_2, \ldots, E_n$ constitute n partition of sample space S, where is $P(E_i) \neq 0$ for $i = 1, 2, 3, \ldots n$ then for any event E define in the sample space S such that $P(E) > 0$. The probability than an event E_k occur given that event E has occured is given by:

$$P(E_k/E) = \frac{P(E_k).P(E/E_k)}{P(E_1)P(E/E_1)+P(E_2).P(E/E_2)+\cdots+P(E_n).P(E/E_n)}$$

$$= \frac{P(E_k).P(E/E_k)}{\sum_{i=1}^{n} P(E_i).P(E/E_i)} \quad \text{---} \ (1)$$



Proof :-

$E = E \cap S$

$E = E \cap (E_1 \cup E_2 \cup \cdots \cup E_n)$

$= (E \cap E_1) \cup (E \cap E_2) \cup \cdots \cup (E \cap E_n)$ [By Distributive law].

as events are mutually exclusive, so:

$$P(E) = P(E \cap E_1) + P(E \cap E_2) + \cdots + P(E \cap E_n) \quad \text{---} \ (2)$$

Scanned with CamScanner

we know that conditional probability is defined as:-

$$P(E/E_1) = \frac{P(E \cap E_1)}{P(E_1)}$$

$$\Rightarrow P(E \cap E_1) = P(E_1).P(E/E_1) \quad \text{---} \quad (3)$$

$P(E_2)$.

So similarly:-

$P(E_r$ $\quad P(E \cap E_2) = P(E_2).P(E/E_2)$.

$\vdots$

$$P(E \cap E_n) = P(E_n).P(E/E_n).$$

Substituting eq'n (3) in (2) we have:

$$P(E) = P(E_1).P(E/E_1) + P(E_2).P(E/E_2) + \cdots + P(E_n).P(E/E$$

which is called law of total probability.

$$P(E) = \sum_{i=1}^{n} P(E_i).P(E/E_i) \quad \text{---} \quad (4)$$

Given that E_k be any other event than:

$$P(E_k/E) = \frac{P(E_k \cap E)}{P(E)} = \frac{P(E_k).P(E/E_k)}{P(E_1).P(E/E_1) + P(E_2).P(E/E_2) + \cdots + P(E_n).P(E/E}$$

$$= \frac{P(E_k).P(E/E_k)}{\sum_{i=1}^{n} P(E_i).P(E/E_i)} \quad \text{Answer}$$

Scanned with CamScanner

Problem:-

Let 10% of male & 2% of female are color blind and each form 50% of total population.

A research ~~working~~ worker studying colorblindness selects a person at random. What is the probability that the so selected is:

a) Male

b) Female

Solution:-

Sex	Color-blind	Not Color-blind	Total
Male	5 (10% of 50)	45	50
Female	1 (2% of 50)	49	50
Total	6	94	100

a) $P(M/C) = \frac{P(M \cap C)}{P(C)} = \frac{5/100}{6/100} = \boxed{\frac{5}{6}}$

a) $P(F/C) = \frac{P(F \cap C)}{P(C)} = \frac{1/100}{6/100} = \boxed{\frac{1}{6}}$

Answer

Scanned with CamScanner

17-Feb-2020.

Lecture # 6

★ Probability Density Function (PDF).

Every thing that can be known about a random variable is contained in it's pdf.

Pdf is a rule that allows us to determine the probability that a particular measurement of a random variable might be within some set of values

★ Random Variable:-

A random variable is the concept of making some measurements.

- The value of measurement need not be the same everytime, so we speak of probabality distribution of random variable.
- we know everything there is to know about random variable on to we know the probability that the value of random variable might be in any specified set.

eg.

sex, age, height,

Scanned with CamScanner

Random Variable:-

A random variable is a function on the sample space & is basically a device for transferring probabilities from complicated sample spaces to simple sample spaces, where the elements are just natural numbers.

- A random variable x is a function whose domain is the sample space and whose range is the set of real numbers. The random number assign a real value (i.e a number) to every outcomes in the sample space.
- The particular values are called realizations and are denoted by 'x'.
- If the realizations are countable, $x_1, x_2, \ldots x_n$ the random variable is said to be **discrete**. [means in whole no.]
- In contrast, if there are infinite, many, uncountable, the random variable is said to be **continuous**. [means not in whole no. in decimal.]

Scanned with CamScanner

- Probability Mass Function. (discrete one).
If x is discrete random variable, the function given by $f(x) = p(x = x)$ for every x within range of x, is called Probability Mass Function or Probability Density Function of x.

- A function can serve as a probability distribution of a discrete random variable x if its values $f(x)$ satisfy the conditions as.

1) $f(x) \geq 0$; for all x within range of x.

2) $\sum_{x} f(x) = 1$; where the summations extends over the values within its domain.

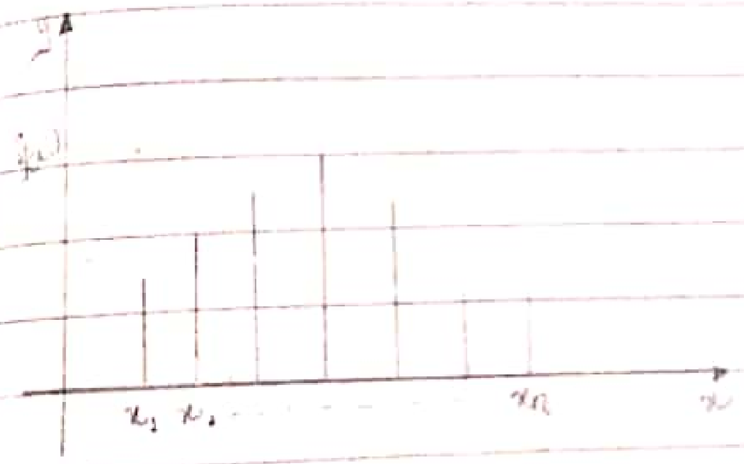
Thus a table, formula or graph representing the probabilities which a random variable x takes is calla probability distribution

Scanned with CamScanner

Table:-

x	x_1	x_2	----	x_n	Total.
$f(x)$	$f(x_1)$	$f(x_2)$	--	$f(x_n)$	$\sum f(x_i) = 1$
$P(x)$	$P(x_1)$	$P(x_2)$	---	$P(x_n)$	$\sum P(x_i) = 1$

Graph:-



Eg.

Distribution of boys in a family having three children. Soln:-

$S \cdot S = n(S) = 2^3 = 8$ cases [children]

[possibilities for a child (each)]

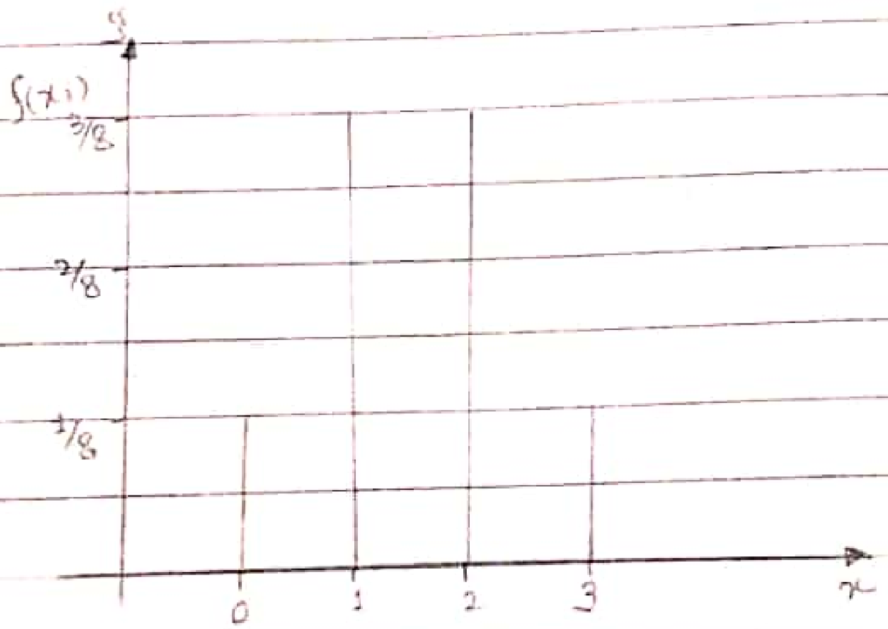
	Children			
Cases.	1	2	3	Count
1	G	G	G	0
2	G	G	B	1
3	G	B	G	1

$\Rightarrow \binom{3}{x} = \frac{3!}{(3-x)!\,x!}$ $\Rightarrow \binom{3}{2} = \frac{3!}{(3-2)!\,2!}$ $\Rightarrow \binom{3}{3}$

$\Rightarrow 3C_2 = 3$

4	G	B	B	2
5	B	G	G	1
6	B	G	B	1
7	B	B	G	2
8	B	B	B	3
				8 cases.

x — C# of boys.	$P(x)$	$P(x) = \binom{3}{x}$
0	$1/8$	
1	$3/8$	
2	$3/8$	
3	$1/8$	
Total.	1	



Probability Histogram.

★ Probability Distribution :-

A random variable 'x' is said to be continuous if there exists a non-negative function f, defined for all $x \in (-\infty, \infty)$ having the property that for any set

ture
= 0,1,2,3.

set of real numbers

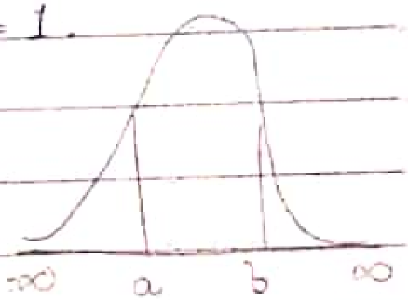
$$P\{x \in A\} = \int_{A} f(x)\,dx.$$

The function f is called the probability density function of random variable x. since x must assign some value, f must satisfy:

$$P\{x \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)\,dx = 1.$$

if $A = \{a, b\}$ then:

$$P\{x \in (a,b)\} = \int_{a}^{b} f(x)\,dx.$$



$$\Rightarrow P\{a \leq x \leq b\} = \int_{a}^{b} f(x)\,dx = F(b) - F(a)$$

$$\Rightarrow P(x < a) = P(x \leq a)$$

$$= F(a) = \int_{-\infty}^{a} f(x)\,dx.$$

• $F(a)$ is called commulative density function or distribution fn.

Scanned with CamScanner

* **Commulative Distributive Function (cdf) :-**

The cdf $F(x)$ of a discrete random function variable x with pmf $P(x)$ or $f(x)$ is defined for every number x by.

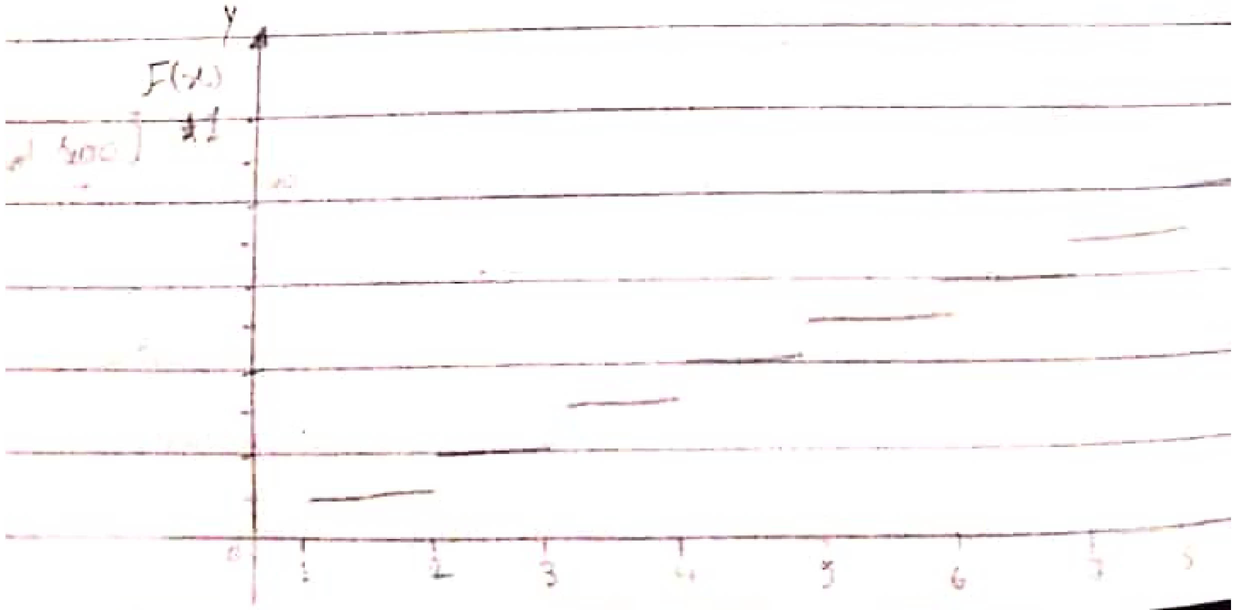
$$F(x) = P(x \le x) = \sum_{x=0}^{x} f(x) = \sum_{x=0}^{x} P(x).$$

For any no. x, $F(x)$ is the probability that the observed value of x will be at most x.

For a Continuous random variable:

$$\boxed{F(a) = P(x<a] = P[x \le a) = \int_{-\infty}^{\infty} f(x)\, dx}$$

For x is a discrete random variable, the graph will have a jump at every possible value of x and will be flat b/w any two possible values of x such a graph is called step function.



Q # 1:

Let x be the random variable with pdf:

$$f(x) = \frac{1}{k}\left(x - \frac{1}{2}\right); \quad k < x < 2$$

Find the value of k so that $f(x)$ be a pdf?

Solution:-

For a pdf:-

$$\therefore \int_k^2 f(x)\,dx = 1$$

$$\Rightarrow \int_k^2 \left(x - \frac{1}{2}\right) dx = k \quad \text{or} \quad \frac{1}{k}\int_k^2 \left(x - \frac{1}{2}\right) dx = 1$$

$$\Rightarrow \frac{1}{k}\left|\frac{x^2}{2} - \frac{1}{2}x\right|_k^2 = 1 \quad \Rightarrow \left[\frac{2^2 - (2)}{2} - \frac{k^2 - k}{2}\right] = 1$$

$$\Rightarrow \frac{2 - k^2 + k}{2} = 1$$

$$\Rightarrow k^2 + k - 2 = 0$$

$$\Rightarrow k^2 + 2k - k - 2 = 0$$

$$\Rightarrow k(k+2) - 1(k+2) = 0$$

$$= (k+2)(k-1)$$

$$k = 1, \quad k = -2$$

Next Class

Mathematical Expectation (Random variable)

Scanned with CamScanner