PROBABILITY AND STATISTICS FOR ENGINEERS

* Data set :- Collected Data (Population) - Ω (capital omega)
* Unit :- Entity of a data (also called as observation) (ω)
* Sample Space :- Selection of observations.
  $\{\omega_1, \omega_2, \ldots \omega_n\} \in \Omega$ $\quad \omega \in \Omega$

* Variable :- A particular feature of any observation, denoted by X.
  $X\ \Omega \rightarrow S$

(A) Qualitative and Quantitave Variables :-

- Qualitative: Variable which can take values that cannot be ordered in a logical or natural way
  e.g. hair colours
- Quantitative: Variables which represent is measurable quantity
  e.g. weight.

(B) Discrete and Continuous Variables :-

- Discrete: Variables that can take finite no. of values
  e.g. eye colour
- Continuous: Variables that can take infinite no. of values
  e.g. weight

(C) Scales :-
used to classify the consideration of variables.

① Nominal Variables :- Values cannot be ordered.
e.g. arranging things according to your preference.

② Ordinal Variables :- Values can be ordered, but difference between the values cann't be interpreted in meaningful way.
e.g. Satisfaction with product (unsatisfied - satisfied - very satisfied)

Theorem 3
The number of permutations of n objects arranged in a circle is $(n-1)!$

Theorem 4
: The no of distinct permutations of n objects, out of which n_1 one ~~kind~~ time, n_2, n_2 times, ---- n_k, r_k time ---- given by $\frac{n!}{n_1!\, n_2! \cdots n_k!}$

Theorem 5
: The no. of ways of partitioning a set of n objects into r cells with n_1 elements ~~in the cell~~ first cell, n_2 elements in the second cell, ------ given by

$$\frac{n!}{n_1!\, n_2! \cdots n_r!} \quad \text{where } n_1+n_2+\cdots+n_r = n$$

Que: How many different letter arrangements can be made from the letters in the word ~~'statistics'~~ "STATISTICS"

$$\Rightarrow \frac{10!}{3!\,3!\,2!} = \frac{10\times9\times8\times7\times\cancel{6}^{3}\times5\times4}{3\times2\times1\times2\times1} = \cancel{75600}, 50400$$

(2) COMBINATORICS $= \frac{13!}{2!\,2!} =$

PROBABILITY :

(1) Random Experiment :
: A experiment that can be repeated any no of times under the same set of conditions and it's outcome is known only after the completion of the experiment, is called a random experiment.
e.g. Tossing of a coin and rolling of a dice.

(2) Simple Events :-
: A possible outcome of a random experiment is called a simple event, it is also called as an elementary event. It is denoted by (ω_i)

(3) Sample Space :
: The set of all possible outcomes of a random experiment is a called a sample space. It is denoted by $(\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\})$

(4) Events :
: Subsets of sample space (Ω) are events. They are denoted by A, B, C, ----

(5) Composite or Complimentary Events :-
: It reffers to the non-occuring of the event. It is denoted by $(\bar{A})$

NOTE : Ω is an event which always occurs and so it is called as sure events or certain event. on the other hand $\phi = \{\ \}$ is an event called as impossible event

(6) Venn Diagram :-
: In venn diagram two or more sets are visualize by circles Overlapping circles implies that both the events have one or more identical simple events, Separated circles shows that

Corollary 4: If $A \subseteq B$, then $P(A) \le P(B)$

$A' \cup (\bar{A} \cap B)'$
$A' \cup (A \cup B')$
$A' \cup B'$

Proof: A and $\bar{A} \cap B$ are disjoint sets

$A \cup (\bar{A} \cap B) = B$

$P(A \cup (\bar{A} \cap B)) = P(B)$

By axiom 3

$P(A) + P(\bar{A} \cap B) = P(B)$

→ Here probabilities get added (i.e. $P(A)$ is getting added to $P(\bar{A} \cap B)$)

$\therefore P(A) \le P(B)$

NOTE: ① $0 \le P(A) \le 1$

② $P(\Omega) = 1$

③ $P(A_1 \cup A_2) = P(A_1) + P(A_2)$; if A_1 & A_2 are disjoint event

④ $P(\phi) = P(\bar{\Omega}) = 0$

⑤ $P(\bar{A}) = 1 - P(A)$

⑥ $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

⑦ If $A \subseteq B$, then $P(A) \le P(B)$

Conditional Probability :-

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

$\frac{1}{2}$ $\frac{1}{4}$

example: tossing of three coins.

HHH
HHT
HTH
HTT
TTT
TTH
THT
THH

E : Getting atleast two heads, $P(E) = 4/8 = 1/2$

F : Getting ~~atleast two~~ First coin as tail $\therefore P(F) = 4/8 = 1/2$

$P(E \cap F) = \frac{1}{8}$

$P(E/F) = \frac{1}{4}$

Definition: Let A be an event, such that probability of A is strictly greater than 0 i.e. $P(A) > 0$, then the conditional probability of event B, given that, event A has already occured is $\boxed{P(B/A) = \frac{P(B \cap A)}{P(A)}}$

Theorem: Multiplication theorem of probability.
For any two arbitrary events A and B, the following equation holds.

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

Theorem: law of total probability.

Assume that $A_1, A_2, \ldots, A_k$ are events such that
$A_1 \cup A_2 \cup \ldots \cup A_k = \Omega$ and $A_i \cap A_j = \phi \;\; \forall\, i \neq j$
$P(A_i) > 0 \;\; \forall\, i$
i.e. $A_1, A_2, \ldots, A_k$ forms a complete decomposition of Ω in Pairwise Disjoint Events.
Then, the probability of an event B is

$$\boxed{\begin{aligned} P(B) &= P(B/A_1)\,P(A_1) + P(B/A_2)\cdot P(A_2) + \cdots + P(B/A_k)\cdot P(A_k) \\ &= \sum_{i=1}^{k} P(B/A_i)\cdot P(A_i) \end{aligned}}$$

Que In a class 35% of students studies science and History, 65% of students studies science, what is the probability of a student studying History given that he or she is already studying science.

→ ~~$P(S) = 0.35$~~, $P(\cancel{H \cup} S) = 0.65$

$P(H \cap S) = 0.35$, $P(S) = 0.65$ $\quad \frac{0.35}{0.65} =$

~~$P(H \cup S) = P(S) + P(H) - P(H \cap S)$~~

~~$0.35 = 0.65 +$~~

$\therefore P(H/S) = \frac{P(H \cap S)}{P(S)} = \frac{0.35}{0.65} = \frac{7}{13}$

Rainbow

(2) CDF of Discrete Random Variable:-

Let, a discrete Random Variable X, then it's a CDF if defined as

$$F(x) = P(X \le x_j)$$
$$= \sum_{i=1}^{k} I_{\{x_i \le x\}}\, p_i$$

where $I_{\{x_i \le x\}} = \begin{cases} 1, & \text{if } x_i \le x \\ 0 & \text{else} \end{cases}$

is called an Indicator function.

• PROPERTIES:-

1) $P(X \le a) = F(a)$
2) $P(X > a) = 1 - P(X \le a)$
3) $P(X < a) = P(X \le a) - P(X = a)$
$\quad = F(a) - P(X = a)$
4) $P(X \ge a) = 1 - P(X < a)$
$\quad = 1 - [F(a) - P(X = a)]$
$\quad = 1 - F(a) + P(X = a)$
5) $P(a \le X \le b) = P(X \le b) - P(X < a)$
$\quad = P(X \le b) - [F(a) - P(X = a)]$
$\quad = P(X \le b) - F(a) + P(X = a)$
6) $P(a < X \le b) = F(b) - F(a)$
7) $P(a < X < b) = F(b) - F(a) - P(X = b)$
8) $P(a \le X < b) = F(b) - F(a) - P(X = b) + P(X = a)$

Rainbow

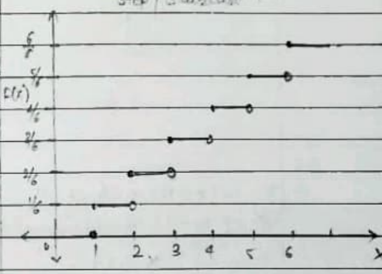
Que. Consider an Experiment of Rolling a Dice.
X: Random Variable as no of Dots upper surface of Dice.
Find PMF and CDF.

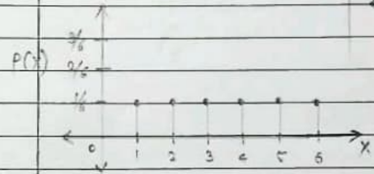
→

X	1	2	3	4	5	6
P(X)	1/6	1/6	1/6	1/6	1/6	1/6

$$F(x) = \begin{cases} 0 & x < 1 \\ 1/6 & x \le 1 \\ 2/6 & x \le 2 \\ 3/6 & x \le 3 \\ 4/6 & x \le 4 \\ 5/6 & x \le 5 \\ 1 & x \le 6 \end{cases}$$

Step / Staircase fn





Q. Consider the following Distribution Table:-

X	0	1	2	3	4
P(X=x)	0	k	2k	2k	

Rainbow

$\therefore \ \underline{K = \frac{1}{10}}$

(ii) $P(X<6) = 0 + \frac{1}{10} + \frac{2}{10} + \frac{2}{10} + \frac{3}{10} + \frac{1}{100}$

$= \frac{81}{100}$

$P(X \geq 6) = 1 - P(X<6)$

$= 1 - \frac{81}{100} = \frac{19}{100}$

$P(0<X<5) = \frac{1}{10} + \frac{2}{10} + \frac{2}{10} + \frac{3}{10} = \frac{8}{10}$

(i) CDF

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{10} & x \leq 1 \\ \frac{3}{10} & x \leq 2 \\ \frac{5}{10} & x \leq 3 \\ \frac{8}{10} & x \leq 4 \\ \frac{81}{100} & x \leq 5 \\ \frac{83}{100} & x \leq 6 \\ 1 & x \leq 7 \end{cases}$$

(iii) Given $P(X \leq c) > \frac{1}{2}$

$P(X \leq 0) = 0, \ P(X \leq 1) = 1, \ P(X \leq 2) = \frac{3}{10}, \ P(X \leq 3) = \frac{5}{10}$

$P(X \leq 4) = \frac{8}{10}$

$\therefore \ \underline{c = 4}$

(iv) $P\left(\frac{1.5<X<4.5}{X>2}\right)$

let $P(1.5<X<4.5) = P(A)$

and $P(X>2) = P(B)$

$\therefore P\left(\frac{A}{B}\right) \ \therefore \ P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{5/10}{7/10} = \frac{5}{7}$

Q. A Random variable has PMF

X	0	1	2	3	4	5	6
P(X=x)	k	3k	5k	7k	9k	11k	13k

Find (i) k (ii) $P(X>2)$ (iii) $P\left(\frac{X>6}{X>2}\right)$

(i) We know that $\sum_{i=1}^{n} P_i = 1$

$\therefore k + 3k + 5k + 7k + 9k + 11k + 13k = 1$

$\therefore 49k = 1 \Rightarrow \boxed{k = \frac{1}{49}}$

(ii) $P(X>2) = 1 - P(X \leq 2)$

$\therefore P(X \leq 2) = \frac{1}{49} + \frac{3}{49} + \frac{5}{49} = \frac{9}{49}$

$\therefore P(X>2) = 1 - \frac{9}{49} = \frac{40}{49}$

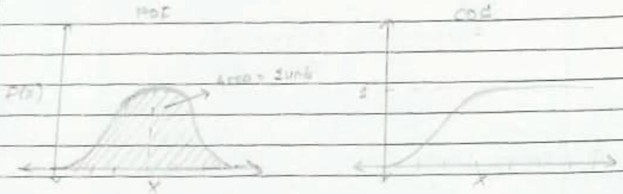
(iii) let $P(X>6) = P(A)$ & $P(X>2) = P(B)$

$\therefore P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)} = \frac{\phi}{40/49} = 0$

It is satisfying 2^{nd} condn also

$\therefore$ It is PDF.

Q. Consider the heights of students of 20 years of age
Graphs for PDF and CDF



• **Theorem:**

Let X be a Continuous Random Variable with CDF $F(x)$. If $x_1 < x_2$, where x_1 and x_2 are unknown constants then:

$$P(x_1 \le X \le x_2) = F(x_2) - F(x_1)$$

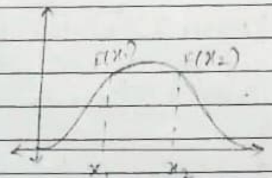
$$P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x)\,dx$$

Proof: Consider the graph of PDF of a Continuous Random Variable X.

$$P(x_1 \le X \le x_2) = F(x_2) - F(x_1)$$

$= (\text{Area under } f(x) \text{ from } -\infty \text{ to } x_2) - (\text{Area under } f(x) \text{ from } -\infty \text{ to } x_1)$

$= (\text{Area under } f(x) \text{ from } x_1 \text{ to } x_2)$



$$\therefore \boxed{P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x)\,dx}$$

Q. Consider a continuous random variable,
X: waiting time of a train
Assume that the train arrives after every 20 minutes.
Find PDF of given information.

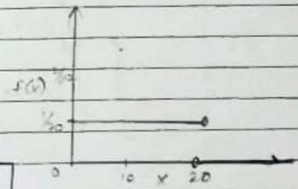
$\rightarrow$ PDF $f(x) = \begin{cases} k & x \in [0,20] \\ 0 & \text{elsewhere.} \end{cases}$

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

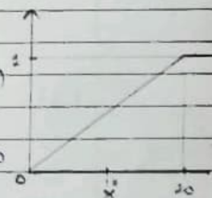
$\therefore 0 + \int_0^{20} k\,dx = 1$

$\therefore kx \Big|_0^{20} = 1$

$\therefore 20k = 1 \Rightarrow \boxed{k = \frac{1}{20}}$



Ans: PDF $f(x) = \begin{cases} 1/20 & x \in [0,20] \\ 0 & \text{elsewhere} \end{cases}$



CDF $F(x) = \begin{cases} \int_{-\infty}^{x} 0\,dt & x \le 0 \\ 0 + \int_0^x \frac{1}{20}\,dt & 0 \le x \le 20 \\ 0 + \int_0^{20} \frac{1}{20}\,dt + 0 & x \ge 20 \end{cases}$

$\therefore F(x) = \begin{cases} 0 & x \le 0 \\ x/20 & 0 < x \le 20 \\ 1 & x \ge 20 \end{cases}$

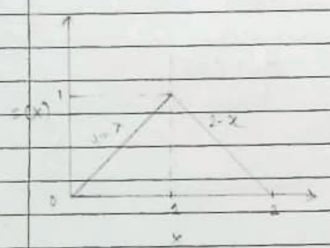
Q. If $f(x) = \begin{cases} x & 0 \le x \le 1 \\ 2-x & 1 \le x < 2 \\ 0 & x \ge 2 \end{cases}$ is a PDF, find its CDF.

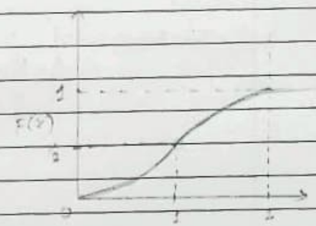
$\rightarrow$ $F(x) = \int_{-\infty}^{\infty} f(t)\,dt$

$$F(x) = \begin{cases} \int_{-\infty}^{x} 0 \cdot dt & x \le 0 \\ \int_{-\infty}^{0} 0 \cdot dt + \int_{0}^{x} t \cdot dt & x \le 1 \\ \int_{-\infty}^{0} 0 \cdot dt + \int_{0}^{1} t \cdot dt + \int_{1}^{x} (2-t)\,dt & x \le 2 \end{cases}$$

$$F(x) = \begin{cases} 0 & x < 0 \\ x^2/2 & x \le 1 \\ \frac{1}{2} + 2x - \frac{x^2}{2} - \left(2 - \frac{1}{2}\right) & x \le 2 \end{cases}$$

$$F(x) = \begin{cases} 0 & x \le 0 \\ x^2/2 & 0 < x \le 1 \\ 2x - \frac{x^2}{2} - 1 & 1 < x \le 2 \\ 1 & x > 2 \end{cases}$$





Q. Find PDF where CDF is

$$F(x) = \begin{cases} 0 & x \le 1 \\ k(x-1)^4, & 1 < x \le 3 \\ 1 & x > 3 \end{cases}$$ Also find k.

$\rightarrow$
$$f(x) = \begin{cases} \frac{d}{dx}F(x) = \frac{d}{dx}(0) = 0 & x \le 1 \\ \frac{d}{dx}F(x) = k\frac{d}{dx}(x-1)^4 = 4k(x-1)^3 & 1 < x \le 3 \\ \frac{d}{dx}F(x) = \frac{d}{dx}(0) = 0 & x > 3 \end{cases}$$

$$\therefore f(x) = \begin{cases} 4k(x-1)^3 & 1 < x \le 3 \\ 0 & \text{elsewhere.} \end{cases}$$

we know, $\int_{-\infty}^{\infty} 4k(x-1)^3\,dx = 1$

$\therefore 0 + \int_{1}^{3} 4k(x-1)^3\,dx = 1$

$\therefore 0 + k(x-1)^4 \Big|_1^3 = 1$

$\therefore k(16) = 1 \Rightarrow \boxed{k = \frac{1}{16}}$

OR, we can find the value of 'k' directly by using continuity at branches
L.H.L = R.H.L = f(x) at branch pt x

$$\therefore f(x) = \begin{cases} \frac{(x-1)^3}{4} & 1 \le x < 3 \\ 0 & \text{elsewhere.} \end{cases}$$

- Continuous Random Variable:-

(1) PDF (Probability Density Function):-
: For a function $f(x)$ to become a probability density function of a continuous random variable X is needs to satisfy the following properties.

(i) $f(x) \geq 0 \quad \forall x$

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

(2) CDF (Cumulative Distribution Function):-
: A random variable, X is said to be continuous if there's a function $f(x)$ st $\forall x$

$$F(x) = \int_{-\infty}^{x} f(t)\,dt \quad \text{holds.}$$

$F(x)$ is the CDF of X and $f(x)$ is the PDF of X.

$\frac{d}{dx}(F(x)) = f(x) \quad \forall x$, that are the points of discontinuity of f.

e.g.

Theorem:
: The probability of a continuous Random Variable at a particular value x is 0 i.e. $P(X = x_0) = 0$

Proof: Consider the interval $(x_0 - \delta, x_0]$ with $\delta > 0$

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

$\therefore P(x_0 - \delta < X \leq x_0) = F(x_0) - F(x_0 - \delta)$

Now, $P(X = x_0) = \lim_{\delta \to 0} P(x_0 - \delta < X \leq x_0)$

$= \lim_{\delta \to 0} (F(x_0) - F(x_0 - \delta)) = F(x_0) - F(x_0)$

$= 0$

NOTE:

$P(a \leq X \leq b) = P(a < X < b) + P(X = b)$

If X is a continuous Random Variable.

then $P(X = b) = 0$

$\therefore P(a < X \leq b) = P(a < X < b)$

and

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

$= P(a \leq X \leq b)$

Q. Prove that

$$f(x) = \begin{cases} x & 0 < x \leq 1 \\ 2 - x & 1 \leq x < 2 \\ 0 & x > 2 \end{cases} \quad \text{is a PDF}$$

→ For a function $f(x)$ to become a probability density f^n (PDF) of a continuous random variable X is needs to satisfy the following following properties

(1) $f(x) \geq 0 \quad \forall x$.

$\therefore$ for $x < 0 \Rightarrow f(x) = 0$

for $0 < x \leq 1 \Rightarrow f(x) = x$

for $1 \leq x < 2 \Rightarrow f(x) = 2 - x$

For $x > 2 \Rightarrow f(x) = 0$.

$\therefore$ It is satisfying 1^{st} cond^n.

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

$\therefore \int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{2} f(x)dx + \int_{2}^{\infty} f(x)dx$

$= \int_{-\infty}^{0} 0 \cdot dx + \int_{0}^{1} x\,dx + \int_{1}^{2} 2 - x\,dx + \int_{2}^{\infty} 0\,dx$

$= 0 + \left.\frac{x^2}{2}\right|_0^1 + \left.2x - \frac{x^2}{2}\right|_1^2 + 0$

$= \frac{1}{2} + \left[(4 - 2) - \left(2 - \frac{1}{2}\right)\right]$

$= \frac{1}{2} + \frac{1}{2} = 1$

Tutorial 2

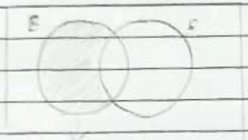
Q.7 let E & F

$P(E \cup F) = 0.7$

$P(E) = 0.5$

$P(F) = 0.3$

E – F



20 × 9

20!/2

$\frac{20!}{17! \, 2!}$

$P(E \cap F^c)$

20 ×

Q.8

→ $P = \{a \mid a \in N, 2 < a < 7, a \text{ is prime}\} = \{3, 5\}$

$Q = \{2b+1 \mid b \text{ is odd number below } 4\} = \{3, 7\}$

$P \cup Q = \{3, 5, 7\}$

$P \cap Q = \{3\}$

$(P \cup Q) \cap (P \cap Q) = \{3\}$

Q.9

Prove that :-

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

$$(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2$$

$$x_1^2 - 2x_1\bar{x} + \bar{x}^2 + x_2 - 2x_2\bar{x} + \bar{x}^2 + \dots + x_n^2 - 2x_n\bar{x} + \bar{x}^2$$

$$(x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2) - 2\bar{x}(x_1 + x_2 + x_3 + \dots + x_n) + n\bar{x}^2$$

$$\sum_{i=1}^{n} x_i^2 - 2\bar{x} \sum x_i + n\bar{x}^2$$

$$\sum_{i=1}^{n} x_i^2 - 2\bar{x} n\bar{x} + n\bar{x}^2$$

$$\sum_{i=1}^{n} x_i^2 + n\bar{x}^2$$

Tut Week 3

Q.1 Flipping coin = {H, T}

Die tossed = {1, 2, 3, 4, 5, 6}

S = {HH, HT, T1, T2, T3, T4, T5, T6}

B = Not possible = { }

as we will not flip a coin if tail occurs

it means if coin flip twice then it must contain one head

Q.2 (a) $5 \times 4 \times 3 \times 2 \times 1 = 5!$

(b) $5! - (4! \times 2!) = 120 - 48 = 72$

Q.3 35

Q.4 $\bar{x} = 23.5$ yrs

$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 23.5 \times 6$

$\frac{x_1 + x_2 + x_3}{3} = 42 \times 3$

$42 \times 3 + (x_4 + x_5 + x_6) = 23.5 \times 6$

$126 + (x_4 + x_5 + x_6) = 15$

Q. Consider the following Distribution Table.

X	0	1	2	3	4	5	6	7
P(X=x)	0	k	2k	2k	3k	k^2	$2k^2$	$7k^2+k$

Find :

(i) k and make its CDF

(ii) $P(X<6)$, $P(X \geq 6)$, $P(0<X<5)$

(iii) If $P(X \leq C) > \frac{1}{2}$ then find min value of c.

(iv) $P\left(\frac{1.5 < X < 4.5}{X > 2}\right)$

⇒ (i) $\sum_{i=1}^{n} p_i = 1$

$\therefore 0 + k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1$

$\therefore 10k^2 + 9k - 1 = 0$

$\therefore k = \frac{1}{10}$, $k = -1$ — not possible.

Random Variables :

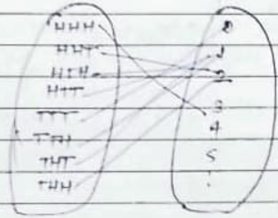
: Let Ω represents the sample space of a random experiment let $\mathbb{R}$ denotes the set of real numbers then a random variable is a function (denoted by X) which assigns each element of sample space one and only one real number i.e. $X : \Omega \longrightarrow \mathbb{R}$

Eg. Ω = Tossing of 3 coins

HHH, HHT, HTH, HTT,

TTT, TTH, THT, THH

Rule : no. of heads



Prob. Distribution Table :-

X (no. of heads)	0	1	2	3
$P(x)$	$1/8$	$3/8$	$3/8$	$1/8$

* Cummulative Distribution Function (CDF) :—

: The CDF of a Random Variable X is defined as

$$F(x) = P(X \le x)$$

$F(x)$ satisfies the following properties :-

① $F(x)$ is a ~~no~~ monotonically non-decreasing function

② $\lim_{x \to -\infty} F(x) = 0$ (lower limit is 0)

③ $\lim_{x \to +\infty} F(x) = 0$ (upper limit is 1)

④ $F(x)$ is continuous from Right

⑤ $0 \le F(x) \le 1 \quad \forall x \in \mathbb{R}$

Que 4 bad apples are mixed with 16 good apples. Find probability distribution table of no. of bad apples when 2 apples are drawn at a random. also find its CDF

Probability Distribution Table :—

X	0	1	2
$P(x)$	$16/20$	$/20$	$\cancel{4}/20$

X	0	1	2
$P(x)$	$\frac{^{16}C_2}{^{20}C_2} = \frac{12}{19}$	$\frac{^4C_1\,{}^{16}C_1}{^{20}C_2} = \frac{32}{95}$	$\frac{^4C_2\,{}^{16}C_0}{^{20}C_2} = \frac{3}{95}$

CDF

$$F(x) = \begin{cases} 0 & x < 0 \\ 92/95 & x \le 1 \\ 1 & x \le 2 \end{cases}$$

Types of Random Variables :—

Two types of random variables :-

(1) Discrete Random Variable :-

: A random variable which takes finite or at max countable number of variables is called Discrete Random Variable.

e.g. No. of heads in toss of two coins

(2) Continuous Random Variable :-

: A random variable which can take infinite no. of values is called continuous Random variable.

eg. Weight of students in class.

• Discrete Random Variable :

(1) Probability Mass Function (PMF) :-

: Let, X be a discrete random variable such that

$P(x) = P(X = x_i) = P_i$ for each $i = 1, 2, \ldots, k$

Then P_i is called PMF if it satisfies the following condn:

(i) $0 \le P_i \le 1$

(ii) $\sum_{i=1}^{k} P_i = 1$

Bayes' Theorem :-

: It gives relation between $P(A/B)$ and $P(B/A)$

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

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

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

$$\boxed{P(A/B) = P(B/A) \times \frac{P(A)}{P(B)}}$$

NOTE :- let $A_1, A_2, \ldots, A_k$ be the events such that

$A_1 \cup A_2 \cup \ldots \cup A_k = \Omega$ and $A_i \cap A_j = \phi \;\forall i \neq j$

$P(A_i) > 0 \quad \forall i$

B is another events,

then $$\boxed{P(A_j/B) = \frac{P(B/A_j)\cdot P(A_j)}{\sum_{j=1}^{k} P(B/A_j)\cdot P(A_j)}}$$

$P(A_j)$ = Prior Probability

$P(B/A_j)$ = Model Probability

$P(A_j/B)$ = Posterior Probability

Independent Events :-

: Two random events A and B are called independent events if ~~Probability of~~ $P(A\cap C) = P(A)\cdot P(B)$ i.e. if the probability of simultaneous occurrence of both the events A and B is taken when it is the product of individual probabilities of A and B.

Independent events are also called as stochastically independent events

Mutually Independent Events :-

: The events $A_1, A_2, \ldots, A_k$ are mutually independently events if for any m events $A_{i_1}, A_{i_2}, \ldots, A_{i_m}$ $(m \leq k)$,

$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_m}) = P(A_{i_1}).P(A_{i_2}) \ldots\ldots P(A_{i_m})$

Que. Consider a bag with 4 balls, the following combination of zeroes and ones are printed on the balls : 110, 101, 011, 000

One ball drawn from the bag, now consider the following events

A_1 : The 1st digit is 1

A_2 : The 2nd digit is 1

A_3 : the 3rd digit is 1

Find ~~the~~ whether the events are independent or not?

→ $A_1 \Rightarrow P(A_1) = \frac{1}{2}$

$A_2 \Rightarrow P(A_2) = \frac{1}{2}$

$A_3 \Rightarrow P(A_3) = \frac{1}{2}$

$P(A_1 \cap A_2) = P(A_1).P(A_2)$
$= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

$P(A_2 \cap A_3) = P(A_2)\,P(A_3)$
$= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

$P(A_1 \cap A_3) = P(A_1).P(A_3)$
$= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

Mutually Indepent Events:-

$P(A_1 \cap A_2 \cap A_3) = 0 \neq P(A_1)\cdot P(A_2)\cdot P(A_3) = \frac{1}{8}$

∴ They are pairwise independent but not mutually Independent.

Ques $X = \{x / x = 3n-1 ;\ n < 3\} = \{2, 5\}$
$Y = \{y / y \text{ is prime no.} < 7\} = \{2, 3, 5\}$
Find $X \cap Y = \{2, 5\}$

E.g. Rolling a dice to get an even no. :-

① Random Exp — Yes
② Sample space - $\{\{1\}, \{2\}, \{3\}, \ldots \{6\}\}$
③ Simple event - $\{1\}$ or $\{2\}$
④ Event — $\{\{2\}, \{4\}, \{6\}\}$
⑤ Complementary Event - $\{\{1\}, \{3\}, \{5\}\}$

• Axiomatic Definition of Probability :-

Axiom 1 :- The every random event A has a probability in $[0, 1]$ i.e. $0 \le P(A) \le 1$

Axiom 2 :- The sure event has a probability 1 i.e. $P(\Omega) = 1$

Axiom 3 :- If A_1 and A_2 are disjoint events then
$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

NOTE : Theorem of ~~editivit~~ additivity of disjoint events:
If $A_1, A_2, \ldots A_k$ are disjoint event then
$$P(A_1 \cup A_2 \ldots \cup A_k) = P(A_1) + P(A_2) + \ldots + P(A_k)$$

Q. Suppose an event is define as the number of points observed on the upper surface of a dice when rolling it. What is the probability of getting a prime number.

~~Prime numb~~ $P(E) = \frac{1}{2}$

How to show using axiom 3?
$$\therefore P(\{2\} \cup P\{3\} \cup P\{5\}) = P(\{2\}) + P(\{3\}) + P(\{5\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

Corollary 1 : The probability of a complementary event of A is
$$P(\bar{A}) = 1 - P(A)$$

Proof :- We know that A and $\bar{A}$ are disjoint events.
and $P(\bar{A}) + P(A) = 1$ — Total probability.
$\therefore P(\bar{A} \cup A) = P(\bar{A}) + P(A) = 1$
$\therefore P(\bar{A}) = 1 - P(A)$

$\therefore$

Proof : By axiom 2, $P(\Omega) = 1$
$\Omega = A \cup \bar{A}$ $\therefore P(\Omega) = P(A \cup \bar{A})$
also, $A \cap \bar{A} = \phi$ $\therefore$ by axiom 3,
$P(A \cup \bar{A}) = P(A) + P(\bar{A}) = 1$
$\therefore P(\bar{A}) = 1 - P(A)$ — proved.

Corollary 2 : The probability of occurrence of an impossible event is 0 i.e. $P(\phi) = P(\bar{\Omega}) = 0$

Proof : $P(\Omega) = 1$ — by axiom 2
by corollary 1 $\Rightarrow P(\bar{\Omega}) = 1 - P(\Omega)$
$\therefore P(\bar{\Omega}) = 1 - 1 = 0$ — hence proved.

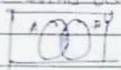
Corollary 3 : Let A_1 and A_2 be not necessarily disjoint events then the probability of occurrence of A_1 or A_2 is $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

Proof :

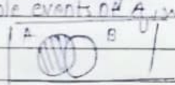
none of the simple events of event A are contain in a sample space of event B

a) $A \cup B$ = The union of events $A \cup B$ is the set of all simple events of A and B which occurrs if atleast one of the simple events of A or B occurs

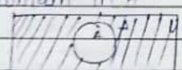
b) $A \cap B$ = The intersection of events is the set of all simple events of A and B which occurs when a simple events occurrs that belongs to both A and B



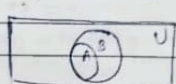
c) $A \setminus B$ (A, but not B) / $A - B$ = The event $A - B$ contains all the simple events of A, which are not contain in B



d) $\overline{A}$ (not A) = It contains all simple events of Ω which are not contain in A



e) $A \subseteq B$ = It contains all the simple events of A, which are also the part of sample space of event B



(7) Disjoint Events :

: Two events A and B are said to ~~dis~~ be disjoint if $A \cap B = \phi$ holds i.e. both events cannot occur simultaneously.

NOTE : The events A and A' are disjoint events.

(8) Mutually disjoint Events :-

: The events $A_1, A_2, \ldots, A_k$ are said to be mutually disjoint events if $A_i \cap A_j = \phi \quad \forall\ i \neq j$

(9) Complete Decomposition :

. The events $A_1, A_2, \ldots, A_k$ forms a complete decomposition of Ω if and only if $A_1 \cup A_2 \cup \ldots \cup A_k = \Omega$ and $A_i \cap A_j = \phi \quad \forall\ i \neq j$

Ques $P = \{a \mid a \text{ is an odd } \cancel{\text{numb}} \text{ prime less than } 7\} = \{3, 5\}$ · $2^2 = 3$

$Q = \{b \mid b \in \mathbb{N};\ 0 \leq b \leq 5\} = \{\cancel{0}, 1, 2, 3, 4\} = 2^4 - 1 = 15$

(a) Find no. of proper subsets of P and Q

→ ~~{3}, only one proper subset.~~ 15+3

(b) $P \cup Q = \{0, 1, 2, 3, 4, 5\}$

(c) $P \setminus Q = \{5\}$

Permutation And Combination

(1) If an operation is performed in n_1 ways and if for each way a second operation can be performed in n_2 ways then the two operations can be performed together in $n_1 \times n_2$ ways.

Q. If a 22 member club meets to elect a chairman and a leader then how many different ways are possible.

⇒ there are 22 ways for elect chairman and then remaining 21 ways are there for elect leader.

$\therefore 22 \times 21 = \underline{462}$ possible ways

(2) If an operation can be performed in n_1 ways and if for each of these a second operation can be performed in n_2 ways and if for each of these a third operation can be performed in n_3 ways (and so on) then all the operations can be performed in $n_1 \times n_2 \times \dots \times n_k$ ways.

Q. How many even 4 digit numbers can be formed from the digits 0, 1, 2, 5, 6, 7, if each of the digit can be used only once.

⇒

~~4! × 2! × 3! × 3! = 72~~

~~5 × 4 × 3 × 2 × 1~~

Case 1:- $\underline{5} \times \underline{4} \times \underline{3} \quad \underline{0} = \cancel{120}\ 60$

Case 2:- $4 \times \cancel{4} \times 3 \times \frac{2}{1} = 96$

$\therefore \cancel{120+36=156} = \underline{156}$

Permutation :- A permutation is an arrangement of all or some part of the set of objects.

Theorem 1 :

The no. of permutation of n objects is $n!$

Theorem 2 :

The no. of permutations of n distinct objects taken r at a time is given by $\boxed{{}^nP_r = \frac{n!}{(n-r)!}}$

Ques: A president and a hunter are to be chosen from a club of 50 members, how many different choices are possible, if

(a) There is no restriction.

(b) If S will serve only if he is a president.

(c) B and D serves together or not at all.

(a)

→ ${}^{50}P_2 = 50 \times 49 = \underline{2450}$

(b)

→

Case 1 : S serves

$\underline{S} \quad \underline{49}$

Case 2 : S does not serve.

${}^{49}P_2 = 49 \times 48 = 2352$

$\therefore 49 + 2352 = \underline{2401}$

(c) Case 1 : B and D serves together

→ $\left.\begin{matrix} \underline{B} & \underline{D} \\ \underline{D} & \underline{B} \end{matrix}\right\} 2$

Case 2 : B and D not serves together.

$\therefore {}^{48}P_2 = 2256$

$\therefore 2 + 2256 = \underline{2258}$

(10) Variance :-

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

sample variance :

$$= \frac{\Sigma(x_i-\bar{x})^2}{(n-1)}$$

Variance doesn't change for scaled data

(11) Standard Deviation :-

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

Q. Consider ~~th~~ 3 students A, B and C arrives at different times in the class to attend their lectures for 5 weeks as shown in the given data, calculate the variance, standard deviation and absolute deviation of the 3 students.

week	1	2	3	4	5
A	0	0	0	0	0
B	-10	+10	-10	+10	-10
C	3	5	6	2	4

(a) Variance of A :

$\bar{x} = 0$

$\therefore \hat{S}^2 = \frac{1}{5} \times 0 = 0$

Variance of B :

$\bar{x} = -2$

$\therefore \hat{S}^2 = \frac{1}{5}[\cancel{144}^{64} + \cancel{64}^{144} + \cancel{144}^{64} + \cancel{64}^{144} + \cancel{144}^{64}]$

$= \cancel{112}\ 96$

Variance of C :

$\bar{x} = 4$

$\hat{S}^2 = \frac{1}{5}[1+1+4+4]$

$= 2$

(b) Standard deviation of A $= 0$

Standard deviation of B $= \cancel{\sqrt{112} = 10.58}$ $\sqrt{96} = 9.798$

Standard deviation of C $= \sqrt{2} = 1.44$

(c) Absolute deviation of A :

$= \frac{1}{5}[0] = 0$

Absolute deviation of B :

$= \frac{1}{5}[\cancel{12} - 8 + 8 - 8 + 8 - 8] = \cancel{10.6}\ 8$

Absolute deviation of C :

$= \frac{1}{5}[1+1+2+2] = 1.2$

Rainbow

Rainbow

④ Quantized : Quantized partitions the data into different promotions.

Let α be a number between 0 and 1 then quantile is given by $\tilde{x}_\alpha = (\alpha \times 100)\%$. which means that quantile divides the data into $(\alpha \times 100)\%$ and $((1-\alpha) \times 100)\%$.

1) If $\alpha = 0.1, 0.2$ and so on ... 0.9, then the quantile is called **Deciles**.

2) If $\alpha = 0.2, 0.4, 0.6, \ldots$ then the quantile is called **Quintile**.

3) If $\alpha = 0.25, 0.50, 0.75, \ldots$ then quantile is called **Quartile**

4) If ~~$\alpha \times 100$ be~~ $(\alpha \cdot 100)$ ~~is integ~~ is an integer then quantile is called **percentile**.

5) quantile can be found as ~~$\tilde{x}_\alpha = \{ x_{(k)}$, if $n\alpha$ is not an integer then choose k as the smallest integer strictly greater than $n\alpha$~~

$$\tilde{x}_\alpha = \begin{cases} x_{(k)}, & \text{if } n\alpha \text{ is not an integer then choose k as the smallest integer strictly greater than } n\alpha \\ \frac{1}{2} x_{(n\alpha)} + x_{(n\alpha+1)}, & \text{if } n\alpha \text{ is an integer} \end{cases}$$

⑤ Mode :-

The mode ($\bar{x}_M$) of ~~m~~ n observations $x_1, x_2, \ldots, x_n$ is the value which occurs the most when compared with other values.

⑥ Absolute Deviation :-

: Consider deviations of n observations around a value A, then the arithmetic mean of all the deviations

$$\boxed{D = \frac{1}{n} \sum_{i=1}^{n} (x_i - A)}$$

The problem with deviation is, it can be positive or negative and hence sum can be very very small or zero and therefore we use modulus and hence

$$\boxed{D = \frac{1}{n} \sum_{i=1}^{n} |x_i - A|}$$

⑦ Absolute Mean Deviation :-

$$D(\bar{x}) = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|$$

⑧ Absolute Median Deviation

$$D(\tilde{x}_{0.5}) = \frac{1}{n} \sum_{i=1}^{n} |x_i - \tilde{x}_{0.5}|$$

NOTE: Note that absolute deviation is minimum when A = median

⑨ Mean squared Error :

$$S^2(A) = \frac{1}{n} \sum_{i=1}^{n} (x_i - A)^2$$

Rainbow

* Measures of central tendency :-

① **Arithmetic mean :-**

Let $x_1, x_2, \ldots, x_n$ be the set of values of a variable. then the arithmetic mean is given by

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

② **Weighted mean :-**

It is used to find mean of grouped data

$$\bar{x} = \frac{n_1 \times m_1 + n_2 \times m_2 + \ldots + \ldots n_k m_k}{n}$$

$$= \sum_{i=1}^{k} \frac{n_i m_i}{n} = \sum_{i=1}^{k} f_i m_i$$

$$= \frac{1}{2} + \frac{9}{2} + \frac{15}{2} + \frac{63}{2} + \frac{86}{2}$$

$$= \underline{\underline{62}}$$

NOTE : The results of mean and weighted mean differ as we use middle class as an approximation of the mean within the scale.

The sum of deviations of each variable around arithmetic mean is zero

$x_1, x_2, x_3, \ldots x_n$

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

Deviations = $\bar{x} - x_1, \bar{x} - x_2, \ldots \bar{x} - x_n$

$(\bar{x} - x_1) + (\bar{x} - x_2) + \ldots + (\bar{x} - x_n)$

$= n\bar{x} - (x_1 + x_2 + x_3 \ldots + x_n) = n\left(\frac{x_1 + x_2 + x_3 \ldots + x_n}{n}\right) - x_1 - x_2 - x_3$

$= 0$

* If

$y_i = a + bx_i$; a, b are constants

then $= a + b\bar{x}$

$x_1, x_2, \ldots x_n$, $\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$

$y_1 = a + bx_1$

$y_2 = a + bx_2$ $\quad \bar{y} = \frac{y_1 + y_2 + y_3 + \ldots + y_n}{n}$

$y_n = a + bx_n$ $\quad = \frac{(a + bx_1) + (a + bx_2) + \ldots + (a + bx_n)}{n}$

$$= \frac{na + bx_1 + bx_2 + \ldots + bx_n}{n}$$

$$= a + b\bar{x}$$

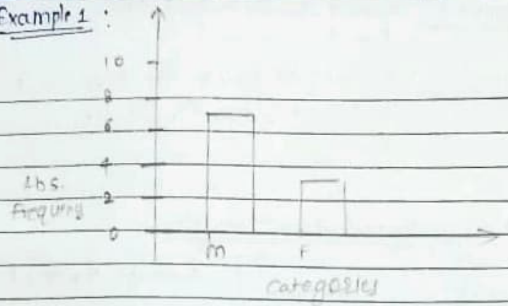
③ **Median :-** $\quad$ 68.5%

Refers to the middle value after arranging the data in ascending order. Denoted by $\tilde{x}_{0.5}$

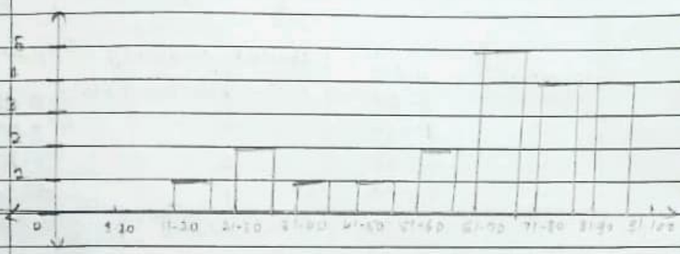
n odd $\quad \tilde{x}_{0.5}$ = middle value

n even $\quad \tilde{x}_{0.5} = \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right)$

Example 1 :



Example 2:

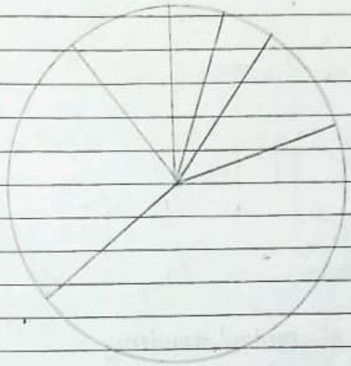


② Pie Chart:

- It represented by a circle partition into segments, where each segment represents a category
- It is used to visualize the absolute frequency of nominal and ordinal values.
- The size of segment depends upon the relative frequency and is determined by the angle given by $f_i \times 360^\circ$

Example 1 :

Category	Abs. freq.	R.F.	angle
m	7	7/10	$\frac{7}{10}$
F	3	3/10	
	10	1	



③ Histogram

- Histogram is used when variables consist of large no. of different values
- It is used to represent the distribution of continuous variables.
- In this we categorized the data in groups and plot the graph by taking bars of each category with height

$h_j = \frac{f_j}{d_j}$; $d_j = e_j - e_{j-1}$

$d_j \to$ width of the class.

ex. 2

Class interval	AF	RF_{f_i}	height (h_j)
0-20	1	1/20	$\frac{1}{20} \times \frac{1}{20} = \frac{1}{400} = 2.5 \times 10^{-3}$
20-40	3	3/20	$\frac{3}{20} \times \frac{1}{20} = \frac{3}{400} = 7.5 \times 10^{-3}$
40-60	3	3/20	$\frac{3}{20} \times \frac{1}{20} = \frac{3}{400} = 7.5 \times 10^{-3}$
61-80	9	9/20	$\frac{9}{20} \times \frac{1}{20} = \frac{9}{400} = 0.0225$
81-100	4	4/20	$\frac{4}{20} \times \frac{1}{20} = \frac{4}{400} = 0.01$
	20	1	20/400

③ Continuous Variable:- Values can be ordered and difference can be interpreted in a meaningful way.
e.g. Natural numbers.

Continuous Variable:-

~~① Ratio~~
① Interval Scale:- In this scale difference between the values can be interpreted but not the ratios.
e.g. temperature.

② Ratio Scale:- In this scale ratio and difference both can be interpreted.
e.g. speed.

③ Absolute Scale:- In this scale values are measured in natural units.
e.g. number of semester studied.

⑤ Grouped / Categorical Data:-
In grouped data, data is available in summarized form.

NOTE:- Any grouped variable which can take only two values are called binary variables.

FREQUENCY:-

* Absolute Frequency:- The number of observations in a particular category (denoted by n_i)

* Relative Frequency:- It gives the comparison between the number of times a number has been repeated to the total frequency of all the number. (denoted by $f_i = \frac{n_i}{n}$)
lies between $0 < f_i < 1$

Example 2: 28, 35, 42, 90, 70, 56, 75, 66, 30, 89, 75, 64, 81, 69, 55, 83, 72, 68, 73, 16,
n=20

• Frequency Distribution Table:-

Example 1:

Category	Absolute frequency n_i	Relative frequency f_i
M	7	7/10
F	3	3/10
	10	1

Example 2:

Class	Absolute frequency n_i	Relative frequency f_i
1-10	0	0.00
11-20	1	0.05
21-30	2	0.10
31-40	1	0.05
41-50	1	0.05
51-60	2	0.1
61-70	5	0.25
71-80	4	0.20
81-90	4	0.20
91-100	0	0.00
	20	1

• Graphical Representation:-

① Bar Chart:-
- It consist of one bar for each category
- Height of the bar is determined by absolute frequency or relative frequency of the respective category on y-axis.
- It is used for nominal and ordinal variables.
- It is used when no. of categories are not too large.