I N D E X

INDEX

UNIT-1 » BASIC PROBABILITY

❖ INTRODUCTION:

- ✓ Probability theory is the branch of mathematics that is concerned with random (or chance) phenomena. It has attracted people to its study both because of its intrinsic interest and its successful applications to many areas within the physical, biological, social sciences, in engineering and in the business world.
- ✓ The words PROBABLE and POSSIBLE CHANCES are quite familiar to us. We use these words when we are sure of the result of certain events. These words convey the sense of uncertainty of occurrence of events.
- ✓ Probability is the word we use to calculate the degree of the certainty of events.
- ✓ There are two types of approaches in the theory of Probability.
  - ➢ Classical Approach – By Blaise Pascal
  - ➢ Axiomatic Approach – By A. Kolmogorov

❖ RANDOM EXPERIMENTS:

- ✓ An experiment is called random experiment if it satisfies the following conditions:
  - ➢ It has more than one possible outcome.
  - ➢ It is not possible to predict the outcome in advance.

❖ SAMPLE SPACE:

- ✓ The set of outcomes is called the sample space of the experiment. It is denoted by "S".
- ✓ If a sample space is in one-one correspondence with a finite set, then it is called a finite sample space. Otherwise it is knowing as an infinite sample space.
  - Examples:
  - ➢ Finite Sample Space: Experiment of tossing a coin twice.

    $\mathrm{S} = \{\mathrm{H}, \mathrm{T}\} \times \{\mathrm{H}, \mathrm{T}\} = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$
  - ➢ Infinite Sample Space: Experiment of tossing a coin until a head comes up for first time.

    $\mathrm{S} = \{\mathrm{H}, \mathrm{TH}, \mathrm{TTH}, \mathrm{TTTH}, \mathrm{TTTTH}, \mathrm{TTTTTH}, \dots\}$

❖ EVENT:

- ✓ A subset of sample space is known as Event. Each member is called Sample Point.
  - **Example:**

    Experiment: Tossing a coin twice. $\mathbf{S} = \{\mathbf{HH}, \mathbf{HT}, \mathbf{TH}, \mathbf{TT}\}$
    Event A: Getting TAIL both times. $\mathbf{A} = \{\mathbf{TT}\}$
    Event B: Getting TAIL exactly once. $\mathbf{B} = \{\mathbf{HT}, \mathbf{TH}\}$

❖ DEFINITIONS:

- ✓ The subset ∅ of a sample space is called "Impossible Events".
- ✓ The subset S(itself) of a sample space is called "Sure/Certain Events".
- ✓ If Subset contains only one element, it is called "Elementary/Simple Events".
- ✓ If Subset contains more than one element, it is called "Compound/Decomposable Events".
- ✓ The set contains all elements other than A is called "Complementary Event" of A. It is denoted by A'.
- ✓ A Union of Events A and B is Union of sets A and B (As per set theory).
- ✓ An Intersection of Events A and B is Intersection of sets A and B (As per set theory).
- ✓ If $A \cap B = \phi$, Events are called Mutually Exclusive Events (Disjoint set).

  Set Notation: $\mathbf{A \cap B = \{ x \mid x \in A \text{ AND } x \in B \}}$

- ✓ If $A \cup B = S$, Events are called Mutually Exhaustive Events.

  Set Notation: $\mathbf{A \cup B = \{ x \mid x \in A \text{ OR } x \in B \}}$

- ✓ If $A \cap B = \phi$ and $A \cup B = S$, Events are called Mutually Exclusive & Exhaustive Events.

❖ DESCRIPTION OF DIFFERENT EVENTS IN WORDS AND SET NOTATION:

No.	Description of Events in Words	Set Notation of Event
1.	A is an event.	A
2.	Event does not occur.	A'
3.	Event B is surely occurs when A occurs.	$B \subset A$
4.	Impossible Events.	Φ
5.	Certain Events.	S (sample space)
6.	Events A and B are Mutually exclusive	$A \cap B = \phi$
7.	Events A and B are Exhaustive	$A \cup B = S$
8.	Among A and B events, Only B occurs	$B - A$ or $B \cap A'$
9.	Among A and B events, Only one event occurs	$(A - B) \cup (B - A)$
10.	Events A and B occur together	$A \cap B$
11.	At least one of the events A and B occurs	$A \cup B$

12.	Among the events A,B,C Only A occurs	$A \cap B' \cap C'$
13.	Among the events A,B,C Only A and B occurs	$A \cap B \cap C'$

METHOD – 1: BASIC EXAMPLES ON SAMPLE SPACES AND EVENTS

H	**1**	Define Mutually Exclusive and Exhaustive events with a suitable example.	
C	**2**	Describe the sample space for the indicated random experiments. **(a)** A coin is tossed 3 times. **(b)** A coin and die is tossed together. **Answer**: $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$ $S = \{H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6\}$	
H	**3**	A balanced coin is tossed thrice. If three tails are obtained, a balance die is rolled. Otherwise the experiment is terminated. Write down the elements of the sample space. **Answer**: $S = \{$ HHH, HHT, HTH, HTT, THH, THT, TTH, TTT1, TTT2, TTT3, TTT4, TTT5, TTT6$\}$	
C	**4**	Let a coin be tossed. If it shows head we draw a ball from a box containing 3 identical red and 4 identical green balls and if it shows a tail, we throw a die. What is the sample space of experiments? **Answer**: $S = \{HR_1, HR_2, HR_3, HG_1, HG_2, HG_3, HG_4, T1, T2, T3, T4, T5, T6\}$	
H	**5**	Four card are labeled with A, B, C and D. We select and two cards at random without replacement. Describe the sample space for the experiments. **Answer**: $S = \{AB, AC, AD, BC, BD, CD\}$	
C	**6**	A coin is tossed 3 times. Give the elements of the following events: Event A: Getting at least two heads Event B: Getting exactly two tails Event C: Getting at most one tail Event D: Getting at least one tail Find $A \cap B$, $C \cap D'$, $A \cup C$, $B \cap C$, $A' \cup C'$. **Answer**: $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$ EVENT A={HHH,HHT,HTH,THH}, EVENT B={HTT,THT,TTH} EVENT C={HHH,HHT,HTH,THH} EVENT D={HHT,HTH,HTT,THH,THT,TTH,TTT}	

T	**7**	There are three identical balls, marked with a, b, c in a box. One ball is picked up from the box at random. The letter on it is noted and the ball is put back in the box, then another ball is picked up from the box and the letter on it is noted. Write the sample space of the experiment. Write down the elements of the following events: Event A: Ball marked “a” is selected exactly once. Event B: Balls selected have same letters marked. Event C: Ball with mark “c” is selected at least once. **Answer**: $S = \{aa, ab, ac, ba, bb, bc, ca, cb, cc\}$, $\text{EVENT A} = \{ab, ac, ba, ca\}$ $\text{EVENT B} = \{aa, bb, cc\}$, $\text{EVENT C} = \{ac, bc, ca, cb, cc\}$	

DEFINITION: PROBABILITY OF AN EVENT

- If a finite sample space associated with a random experiments has "$\mathbf{n}$" equally likely (Equiprobable) outcomes (elements) and of these "$\mathbf{r}$“ ($\mathbf{0} \le \mathbf{r} \le \mathbf{n}$) outcomes are favorable for the occurrence of an event **A**, then probability of **A** is defined as follow.

$$\mathbf{P}(A) = \frac{\text{favorable outcomes}}{\text{total outcomes}} = \frac{r}{n}$$

DEFINITION: EQUIPROBABLE EVENTS

- Let $U = \{x_1, x_2, .., x_n\}$ be a finite sample space. If $P(\{x_1\}) = P(\{x_2\}) = P(\{x_3\}) = \cdots = P(\{x_n\})$, then the elementary events $\{x_1\}, \{x_2\}, \{x_3\}, \dots, \{x_n\}$ are called Equiprobable Events.

RESULTS:

- For the Impossible Event $P(\phi) = 0$.
- Complementation Rule: For every Event A, $P(A') = 1 - P(A)$.
- If $A \subset B$, than $P(B - A) = P(B) - P(A)$ and $P(A) \le P(B)$
- For every event A, $0 \le P(A) \le 1$.
- Let S be sample space and A, B and C be any events in S, then

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

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

(3). $P(A \cap B') = P(A) - P(A \cap B)$

(4). $P(A' \cap B) = P(B) - P(A \cap B)$

(5). $P(A' \cap B') = P(A \cup B)' = 1 - P(A \cup B)$ (De Morgan’s Rule)

(6). $P(A' \cup B') = P(A \cap B)' = 1 - P(A \cap B)$ (De Morgan’s Rule)

❖ PERMUTATION:

✓ Suppose that we are given ‘n’ distinct objects and wish to arrange ‘r’ of these objects in a line. Since there are ‘n’ ways of choosing the 1st object, after this is done ‘n-1’ ways of choosing the 2nd object and finally n-r+1 ways of choosing the r^{th} object, it follows by the fundamental principle of counting that the number of different arrangement (or PERMUTATIONS) is given by

$$\mathbf{nP_r = n\,(n-1)\,(n-2)\ldots\ldots\ldots(n-r+1) = \frac{n!}{(n-r)!}}$$

- **Note:**

(1). Suppose that a set consists of ‘n’ objects of which n_1 are of one type, n_2 are of second type, … …, and n_k are of k^{th} type. Here $n = n_1 + n_2 + \cdots + n_k$. Then the number of different permutations of the objects is

$$\frac{n!}{n_1!\ \ n_2!\ \ \ldots\ \ n_k!}$$

- **Example:** A number of different permutations of the letters of the word MISSISSIPPI is

$$\frac{11!}{1!\ \ 4!\ \ \ 4!\ \ 2!} = 34650$$

(2). If ‘r’ objects are to be arranged out of ‘n’ objects and if repetition of a object is allowed then the total number of permutations is

$$n^r$$

- **Example:** Different numbers of three digits can be formed from the digits 4, 5, 6, 7, 8 is

$$5^3 = 125$$

❖ COMBINATION:

✓ In a permutation we are interested in the order of arrangement of the objects. For example, ABC is a different permutation from BCA. In many problems, however, we are interested only in selecting or choosing objects without regard to order. Such selections are called combination.

✓ The total number of combination (selections) of ‘r’ objects selected from ‘n’ objects is denoted and defined by

$$\mathbf{n\,C_r = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}}$$

- **Examples:**

(1). The number of ways in which 3 card can be chosen from 8 cards is

$$\binom{8}{3} = \frac{8!}{3!\,(8-3)!} = \frac{8\times 7\times 6}{3\times 2\times 1} = 56$$

(2). A club has 10 male and 8 female members. A committee composed of 3 men and 4 women is formed. In how many ways this be done?

$$\binom{10}{3}\binom{8}{4} = 120 \times 70 = 8400$$

(3). Out of 6 boys and 4 girls in how many ways a committee of five members can be formed in which there are at most 2 girls are excluded?

$$\binom{4}{2}\binom{6}{3}+\binom{4}{1}\binom{6}{4}+\binom{4}{0}\binom{6}{5}=120+60+6=186$$

METHOD – 2: PROBABILITY OF EVENTS

H	**1**	If probability of event A is $9/10$, what is the probability of the event "not A". **Answer**: **0.1**	
C	**2**	If A and B are two mutually exclusive events with $P(A)=0.30, P(B)=0.45$. Find the probability of A', $A\cap B$, $A\cup B$, $A'\cap B'$. **Answer**: **0.7 , 0 , 0.75, 0.25**	
T	**3**	The probability that a student passes a physic test is $2/3$ and the probability that he passes both physic and English tests is $14/45$. The probability that he passes at least one test is $4/5$, what is the probability that he passes the English test? **Answer**: **$4/9$**	
C	**4**	A fair coin is tossed twice. Find the probability of **(a)** Getting H exactly once. **(b)** Getting T at least once. **Answer**: **$0.5, 0.75$**	
C	**5**	Two unbiased dice are tossed simultaneously. Find the probability that sum of numbers on the upper face of dice is 9 or 12. **Answer**: **0.14**	
H	**6**	One card is drawn at random from a well shuffled pack of 52 cards. Calculate the probability that the card will be **(a)** An Ace **(b)** A card of black color **(c)** A diamond **(d)** Not an ace. **Answer**: **$0.077, 0.5$, 0.25 , 0.92**	
C	**7**	Four cards are drawn from the pack of cards. Find the probability that **(a)** all are diamonds **(b)** there is one card of each suit **(c)** there are two spades and two hearts. [Hint : Spades ♠ , Club ♣, Diamond ♦, Heart ♥] **Answer**: **$0.0026, 0.1055, 0.0225$**	

H	**8**	Consider a poker hand of five cards. Find the probability of getting four of a kind (i.e., four cards of the same face value) assuming the five cards are chosen at random. **Answer**: $\frac{1}{4165}$	
C	**9**	A box contains 5 red, 6 white and 2 black balls. The balls are identical in all respect other than color. **(a)** One ball is drawn at random from the box. Find the probability that the selected ball is black. **(b)** Two balls are drawn at random from the box. Find the probability that one ball is white and one is red. **Answer**: $2/13, 5/13$	
C	**10**	A room has three lamp sockets. From a collection of 10 light bulbs of which only 6 are good. A person selects 3 at random and puts them in the socket. What is the probability that the room will have light? **Answer**: $\frac{29}{30}$	
C	**11**	Four letters of the words THURSDAY are arranged in all possible ways. Find the probability that the word formed is HURT. **Answer**: $\frac{1}{1680}$	
T	**12**	Find the probability that there will be 5 Sundays in the month of July. **Answer**: 0.43	

❖ CONDITIONAL PROBABILITY:

- ✓ Let S be a sample space and A and B be any two events in S. Then the probability of the occurrence of event A when it is given that B has already occurred is define as

$$P(A/B) = \frac{P(A \cap B)}{P(B)}; P(B) > 0$$

- ✓ Which is known as conditional probability of the event A relative to event B.
- ✓ Similarly, the conditional probability of the event B relative to event A is

$$P(B/A) = \frac{P(B \cap A)}{P(A)}; P(A) > 0$$

- **Properties:**

➢ Let A_1, A_2 and B be any three events of a sample space S,then

$$P(A_1 \cup A_2/B) = P(A_1/B) + P(A_2/B) - P(A_1 \cap A_2/B); P(B) > 0.$$

- Let A and B be any two events of a sample space S, then

$$P(A'/B) = 1 - P(A/B); P(B) > 0.$$

❖ THEOREM (MULTIPLICATION RULE):

- ✓ Let S be a sample space and A and B be any two events in **S**, then

$$P(A \cap B) = P(A) \cdot P(B/A); P(A) > 0 \text{ or } P(A \cap B) = P(B) \cdot P(A/B); P(B) > 0.$$

- **Corollary:**
- Let S be a sample space and A, B and C be three events in S,then

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

❖ INDEPENDENT EVENTS:

- ✓ Let A and B be any two events of a sample space S, then A and B are called independent events if $P(A \cap B) = P(A) \cdot P(B)$.
- ✓ It also means that, $P(A/B) = P(A)$ and $P(B/A) = P(B)$.
- ✓ This means that the probability of A does not depend on the occurrence or nonoccurrence of B, and conversely.

- **Remarks:**
- Let A , B and C are said to be Mutually independent, if

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

(2) $P(B \cap C) = P(B) \cdot P(C)$

(3) $P(C \cap A) = P(C) \cdot P(A)$

(4) $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$

- Let A , B and C are said to be Pairwise independent, if

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

(2) $P(B \cap C) = P(B) \cdot P(C)$

(3) $P(C \cap A) = P(C) \cdot P(A)$

METHOD – 3: CONDITIONAL PROBABILITY & INDEPENDENT EVENT

C	**1**	For two independent events A & B if $P(A) = 0.3$, $P(A \cup B) = 0.6$, find $P(B)$. Answer: $\mathbf{0.429}$	
H	**2**	If A,B are independent events and $P(A) = 1/4, P(B) = 2/3$. Find $P(A \cup B)$. **Answer**: $\mathbf{0.75}$	

H	**3**	If $P(A) = 1/3, P(B) = 3/4$ and $P(A \cup B) = 11/12$. Find $P(A/B)$. **Answer**: $\mathbf{2/9}$	
H	**4**	If A and B are independent events, with $P(A) = 3/8, P(B) = 7/8$. Find $P(A \cup B)$, $P(A/B)$ and $P(B/A)$. **Answer**: $\mathbf{59/64, 3/8, 7/8}$	
C	**5**	If $P(A) = 1/3, P(B') = 1/4$ and $P(A \cap B) = 1/6$, then find $P(A \cup B)$, $P(A' \cap B')$ and $P(A'/B')$. **Answer**: $\mathbf{11/12, 1/12, 1/3}$	
C	**6**	A person is known to hit the target in 3 out of 4 shots, whereas another person is known to hit the target in 2 out of 3 shots. What is the probability that the target will be hit? **Answer**: $\mathbf{11/12}$	
T	**7**	Let S be square $0 \le x \le 1$, $0 \le y \le 1$ in plane. Consider the uniform probability space on square. Show that A and B are independent events if $A: \left\{(x,y): 0 \le x \le \frac{1}{2}, 0 \le y \le 1\right\}$ & $B: \left\{(x,y): 0 \le x \le 1,\ 0 \le y \le \frac{1}{4}\right\}$.	
H	**8**	In a group of 200 students 40 are taking English,50 are taking mathematics,and 12 are taking both. **(a)** If a student is selected at random, what is the probability that the student is taking English? **(b)** A student is selected at random from those taking mathematics. What is the probability that the student is taking English? **(c)** A student is selected at random from those taking English, what is the probability that the student is taking mathematics? **Answer**: $\mathbf{0.20}$, $\mathbf{0.24}$, $\mathbf{0.3}$	
C	**9**	In a box, 100 bulbs are supplied out of which 10 bulbs have defects of type A, 5 bulbs have defects of type B and 2 bulbs have defects of both the type. Find the probability that **(a)** a bulb to be drawn at random has a B type defect under the condition that it has an A type defect, and **(b)** a bulb to be drawn at random has no B type defect under the condition that it has no A type defect. **Answer**: $\mathbf{0.2, 0.9667}$	

C	**10**	From a bag containing 4 white and 6 black balls, two balls are drawn at random. If the balls are drawn one after the other without replacements, find the probability that first ball is white and second is black. **Answer**: $4/15$	

❖ TOTAL PROBABILITY:

- ✓ If B_1 and B_2 are two mutually exclusive and exhaustive events of sample space S and $P(B_1), P(B_2) \neq 0$, then for any event A,

$$\mathbf{P(A) = P(B_1) \cdot P(A/B_1) + P(B_2) \cdot P(A/B_2)}$$

- **Corollary**: If B_1, B_2 and B_3 are mutually exclusive and exhaustive events and $(B_1), P(B_2), P(B_3) \neq 0$, then for any event A.

$$P(A) = P(B_1) \cdot P(A/B_1) + P(B_2) \cdot P(A/B_2) + +P(B_3) \cdot P(A/B_3)$$

❖ BAYES' THEOREM:

- ✓ Let $B_1, B_2, B_3 \dots, B_n$ be an n-mutually exclusive and exhaustive events of a sample space S and let A be any event such that $P(A) \neq 0$, then

$$\mathbf{P(B_i/A) = \frac{P(B_i) \cdot P(A/B_i)}{P(B_1) \cdot P(A/B_1) + P(B_2) \cdot P(A/B_2) + \cdots + P(B_n) \cdot P(A/B_n)}}$$

METHOD – 4: TOTAL PROBABILITY AND BAYE'S THEOREM

C	**1**	A box contains 3 blue and 2 red marbles while another box contains 2 blue and 5 red marbles. A marble drawn at random from one of the boxes. Find the probability that it is a blue marble. **Answer**: $31/70$	
T	**2**	An urn contains 10 white and 3 black balls, while another urn contains 3 white and 5 black balls. Two balls are drawn from the first urn and put into the second urn and then a ball is drawn from the later. What is the probability that it is a white ball? **Answer**: $59/130$	
C	**3**	In a certain assembly plant, three machines, B1, B2 and B3, make 30%, 45% and 25%, respectively, of the products. It is known form the past experience that 2%, 3% and 2% of the products made by each machine respectively, are defective. Now, suppose that a finished product is randomly selected. What is the probability that it is defective? **Answer**: 0.0245	

T	**4**	Suppose that the population of a certain city is 40% male and 60% female. Suppose also that 50% of the males and 30% of the females smoke. Find the probability that a smoker is male. **Answer**: **10/19**	
C	**5**	State Bayes' theorem. A microchip company has two machines that produce the chips. Machine-I produces 65% of the chips, but 5% of its chips are defective. Machine-II produces 35% of the chips, but 15% of its chips are defective. A chip is selected at random and found to be defective. What is the probability that it came from Machine-I? **Answer**: $\mathbf{0.3824}$	
C	**6**	State Bayes' theorem. In a bolt factory, three machines A, B and C manufacture 25%, 35% and 40% of the total product respectively. Out Of these outputs 5%, 4% and 2% respectively, are defective bolts. A bolt is picked up at random and found to be defective. What are the Probabilities that it was manufactured by machine A, B and C? **Answer**: $\mathbf{0.3623, 0.4057, 0.2318}$	
H	**7**	A company has two plants to manufacture hydraulic machine. Plat I manufacture 70% of the hydraulic machines and plant II manufactures 30%. At plant I, 80% of hydraulic machines are rated standard quality and at plant II, 90% of hydraulic machine are rated standard quality. A machine is picked up at random and is found to be of standard quality. What is the chance that it has come from plant I? **Answer**: $\mathbf{0.6747}$	
C	**8**	There are two boxes A and B containing 4 white, 3 red and 3 white,7 red balls respectively. A box is chosen at random and a ball is drawn from it, if the ball is white, find the probability that it is from box A. **Answer**: $\mathbf{\frac{40}{61}}$	
T	**9**	If proposed medical screening on a population, the probability that the test correctly identifies someone with illness as positive is 0.99 and the probability that test correctly identifies someone without illness as negative is 0.95. The incident of illness in general population is 0.0001. You take the test, the result is positive then what is the probability that you have illness? **Answer**: $\mathbf{0.002}$	

H	**10**	In producing screws, let A mean “screw too slim” and B “screw too short”. Let $p(A) = 0.1$ and let the conditional probability that a slim screw is also too small be $P(B/A) = 0.2$. What is the probability of screw, we pick randomly from the lot produced, will be both too slim and too short? **Answer**: $\mathbf{0.02}$	
H	**11**	Three boxes contain $10\%, 20\%,$ and 30% of defective finger joints. A finger joint is selected at random which is defective. Determine probability that it comes from i)1st box ii) 2nd box iii) 3rd box **Answer**: $\mathbf{\frac{1}{6}, \frac{1}{3}, \frac{1}{2}}$	
T	**12**	Suppose there are three chests each having two drawers. The first chest has a gold coin in each drawer, the second chest has a gold coin in one drawer and a silver coin in the other drawer and the third chest has a silver coin in each drawer. A chest is chosen at random and a drawer opened. If the drawer contains a gold coin, what is the probability that the other drawer also contains a gold coin? **Answer**: $\mathbf{2/3}$	

❖ RANDOM VARIABLE:

- ✓ An experiment, in which we know all the possible outcomes in advance but which of them will occur is known only after the experiment is performed, is called a Random Variable.

❖ PROBABILITY DISTRIBUTION OF RANDOM VARIABLE:

- ✓ Probability distribution of random variable is the set of its possible values together with their respective probabilities. It means,

X	x_1	x_2	x_3	…	…	x_n
P(X)	$p(x_1)$	$p(x_2)$	$p(x_3)$	…	…	$p(x_n)$

; Where $p(x_i) \geq 0$ for all i and $\sum_{i=1} p(x_i) = 1$.

- **Example:** Two balanced coins are tossed, find the probability distribution for heads.

  Sample space $= \{HH, HT, TH, TT\}$

  $P(X = 0) = P(\text{no head}) = \frac{1}{4} = 0.25$

  $P(X = 1) = P(\text{one head}) = \frac{2}{4} = \frac{1}{2} = 0.5$

  $P(X = 2) = P(\text{two heads}) = \frac{1}{4} = 0.25$

- Probability distribution is as follow:

X	0	1	2
P(X)	0.25	0.50	0.25

TYPES OF RANDOM VARIABLES:

(1). Discrete Random Variable

(2). Continuous Random Variable

DISCRETE RANDOM VARIABLE:

- ✓ A random variable, which can take only finite, countable, or isolated values in a given interval, is called discrete random variable.
- ✓ i.e. A random variable is one, which can assume any of a set of possible values which can be counted or listed.
- ✓ A discrete random variable is a random variable with a finite (or countably infinite) range.
- ✓ For example, the numbers of heads in tossing coins, the number of auto passengers can take on only the values 1 , 2 , 3 and so on.
  - **Note:** Discrete random variables can be measured exactly.

CONTINUOUS RANDOM VARIABLE:

- ✓ A random variable, which can take all possible values that are infinite in a given interval, is called Continuous random variable.
- ✓ i.e. a continuous random variable is one, which can assume any of infinite spectrum of different values across an interval which cannot be counted or listed.
- ✓ For example, measuring the height of a student selected at random, finding the average life of a brand X tire etc.
  - **Note:** Continuous random variables cannot be measured exactly.

PROBABILITY FUNCTION:

- ✓ If for random variable X, the real valued function f(x) is such that $P(X = x) = f(x)$, then f(x) is called Probability function of random variable X.
- ✓ Probability function f(x) gives the measures of probability for different values of X say $x_1, x_2, \dots, x_n$.

PROBABILITY MASS FUNCTION:

- ✓ If X is a discrete random variable then its probability function p(x) is discrete probability function. It is also called probability mass function.

- **Properties:**
- ➢ $P(X = x_i) = p(x_i)$
- ➢ $p(x) \geq 0$ and $\sum_{i=1}^{n} p(x_i) = 1$

❖ PROBABILITY DENSITY FUNCTION:

✓ If X is a continuous random variable then its probability function f(x) is called continuous probability function OR probability density function.

- **Properties:**
- ➢ $P(a < x < b) = \int_a^b f(x)\,dx$
- ➢ $f(x) \geq 0$
- ➢ $\int_{-\infty}^{\infty} f(x)\,dx = 1$

❖ MATHEMATICAL EXPECTATION:

✓ If X is a discrete random variable having various possible values $x_1, x_2, \dots, x_n$ & if f(x) is the probability function, the mathematical Expectation of X is defined & denoted by

$$\mathbf{E(X) = \sum_{i=1}^{n} x_i \cdot f(x_i) = \sum_{i=1}^{n} x_i \cdot p(x_i) = \sum_{i=1}^{n} x_i \cdot p_i}$$

✓ If X is, a continuous random variable having probability density function f(x), expectation of X is defined as

$$\mathbf{E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx}$$

✓ E(X) is also called the mean value of the probability distribution of x and is denoted by μ.

- **Properties:**
- ➢ Expected value of constant term is constant. i.e. $\mathbf{E(c) = c}$
- ➢ If c is constant, then $\mathbf{E(cX) = c \cdot E(X)}$
- ➢ $\mathbf{E(X^2) = \sum_{i=1}^{n} x_i^2 \cdot p_i}$ (PMF)
- ➢ $\mathbf{E(X^2)} = \int_{-\infty}^{\infty} x^2 f(x)\,dx$ (PDF)
- ➢ If a and b are constants, then $\mathbf{E(aX \pm b) = aE(X) \pm b}$
- ➢ If a , b and c are constants, then $\mathbf{E\left(\frac{aX+b}{c}\right) = \frac{1}{c}[aE(X) + b]}$
- ➢ If X and Y are two random variables , then $\mathbf{E(X + Y) = E(X) + E(Y)}$
- ➢ If X and Y are two independent random variable, then $\mathbf{E(X \cdot Y) = E(X) \cdot E(Y)}$

❖ VARIANCE OF A RANDOM VARIABLE:

- ✓ Variance is a characteristic of random variable X and it is used to measure dispersion (or variation) of X.
- ✓ If X is a discrete random variable (or continuous random variable) with probability mass function f(x) (or probability density function), then expected value of $[X - E(X)]^2$ is called the variance of X and it is denoted by V(X).

$$\mathbf{V(X) = E(X^2) - [E(X)]^2}$$

- **Properties:**
  - ➢ $\mathbf{V(c) = 0}$, Where c is a constant.
  - ➢ $\mathbf{V(cX) = c^2\,V(X)}$, where c is a constant.
  - ➢ $\mathbf{V(X + c) = V(X)}$, Where c is a constant.
  - ➢ If a and b are constants , then $\mathbf{V(aX + b) = a^2 V(X)}$
  - ➢ If X and Y are the independent random variables, then
  - ➢ $\mathbf{V(X + Y) = V(X) + V(Y)}$

❖ STANDARD DEVIATION OF RANDOM VARIABLE:

- ✓ The positive square root of V(X) (Variance of X) is called standard deviation of random variable X and is denoted by σ. i.e. $\boldsymbol{\sigma} = \sqrt{\mathbf{V(X)}}$
  - **Note:** $\boldsymbol{\sigma^2}$ is called variance of $\mathbf{V(X)}$.

❖ CUMULATIVE DISTRIBUTION FUNCTION:

- ✓ An alternate method for describing a random variable's probability distribution is with cumulative probabilities such as $P(X \le x)$.
  - ➢ A Cumulative Distribution Function of a discrete random variable X, denoted as F(X), is

$$\mathbf{F(x) = P(X \le x) = \sum_{x_i \le x} P(x_i)}$$

  - ➢ A Cumulative Distribution Function of a continuous random variable X is

$$\mathbf{F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt\,, -\infty < x < \infty}$$

- **Example:** Determine the probability mass function of X from the following Cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < -2 \\ 0.2 & -2 \le x < 0 \\ 0.7 & 0 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$

- **Answer:** The probability mass function of X is

$$P(-2) = 0.2 - 0 = 0.2, \qquad P(0) = 0.7 - 0.2 = 0.5, \qquad P(2) = 1 - 0.7 = 0.3$$

- Then probability distribution is:

X	-2	0	2
P(X)	0.2	0.5	0.3

- **Properties:**
- $F(-\infty) = 0, F(+\infty) = 1$ and $0 \le F(x) \le 1$
- F is a non-decreasing function, i.e. if $x_1 \le x_2$, then $F(x_1) \le F(x_2)$.
- If $F(x_0) = 0$, then $F(x) = 0$ for every $x \le x_0$.
- $P\{X > x\} = 1 - F(x)$.
- $P(\{x_1 < X \le x_2\}) = F(x_2) - F(x_1)$.

METHOD – 5: RANDOM VARIABLES AND PROBABILITY FUNCTIONS

<table>
<tr><td>C</td><td>1</td><td>Which of the following functions are probability function?
(a) $f(x) = \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{1-x}$; $x = 0,1$ (b) $f(x) = \left(-\frac{1}{2}\right)^x$, $x = 0,1,2$
Answer: YES, NO</td><td></td></tr>
<tr><td>H</td><td>2</td><td>Which of the following functions are probability function?
<table><tr><td>X</td><td>-1</td><td>0</td><td>1</td></tr><tr><td>f(x)</td><td>0.5</td><td>0.8</td><td>-3</td></tr></table>
<table><tr><td>X</td><td>1</td><td>2</td><td>3</td></tr><tr><td>f(x)</td><td>0.3</td><td>0.5</td><td>0.2</td></tr></table>
Answer: NO , YES</td><td></td></tr>
<tr><td>C</td><td>3</td><td>For the probability function $f(x) = \frac{k}{1+x^2}, -\infty < x < \infty$, find k .
Answer: $1/\pi$</td><td></td></tr>
<tr><td>C</td><td>4</td><td>The probability function of a random variable X is $p(x) = \frac{2x+1}{48}$,
$x = 1,2,3,4,5,6$. Verify whether p(x) is probability function Also find E(X).
Answer: YES, 4. 23</td><td></td></tr>
</table>

H **5** Find the expected value of a random variable X having the following probability distribution.

X	-5	-1	0	1	5	8
$P(X = x)$	0.12	0.16	0.28	0.22	0.12	0.1

Answer: $\mathbf{0.8600}$

C **6** The following table gives the probabilities that a certain computer will malfunction 0,1,2,3,4,5, or 6 times on any one day:

NO. of malfunction x:	0	1	2	3	4	5	6
Probability f(x)	0.17	0.29	0.27	0.16	0.07	0.03	0.01

Find mean and standard deviation of this probability distribution.

Answer: $\mathbf{mean = E(X) = 1.8, \sigma = \sqrt{1.8}}$

H **7** The probability distribution of a random variable x is as follows. Find p, E(x).

X	0	1	2	3	4	5
p(x)	P	1/5	1/10	P	1/20	1/20

Answer: $\mathbf{p = 0.30, E(x) = 1.7500}$

C **8** A random variable X has the following function.

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

Find the value of k and then evaluate $P(X < 6), P(X \geq 6)$ and $P(0 < x < 5)$.

Answer: $\mathbf{0.8100, 0.1900, 0.8000}$

H **9** Find $P(2 \leq x < 4), p(X > 2)$ if a random variable X has the p.m.f given by:

X	1	2	3	4
P(X)	0.1	0.2	0.5	0.2

Answer: $\mathbf{0.7, 0.7}$

<table>
<tr><td>C</td><td>10</td><td>A machine produces on average of 500 items during first week of the month & average of 400 items during the last week of the month. The probability for these being 0.68 and 0.32. Determine the expected value of the production.
Answer: $\mathbf{468}$</td><td></td></tr>
<tr><td>H</td><td>11</td><td>In a business, the probability that a trader can get profit of Rs. 5000 is 0.4 and probability for loss of Rs. 2000 is 0.6. Find his expected gain or loss.
Answer: $\mathbf{800}$</td><td></td></tr>
<tr><td>C</td><td>12</td><td>Three balanced coins are tossed, find the mathematical expectation of tails.
Answer: $\mathbf{1.5}$</td><td></td></tr>
<tr><td>C</td><td>13</td><td>There are 8 apples in a box, of which 2 are rotten. A person selects 3 Apples at random from it. Find the expected value of the rotten Apples.
Answer: $\mathbf{0.75}$</td><td></td></tr>
<tr><td>T</td><td>14</td><td>There are 3 red and 2 white balls in a box and 2 balls are taken at random from it. A person gets Rs. 20 for each red ball and Rs. 10 for each white ball. Find his expected gain.
Answer: $\mathbf{32}$</td><td></td></tr>
<tr><td>H</td><td>15</td><td>There are 10 bulbs in a box, out of which 4 are defectives. If 3 bulbs are taken at random, find the expected number of defective bulbs.
Answer: $\mathbf{1.2}$</td><td></td></tr>
<tr><td>H</td><td>16</td><td>(a) A contestant tosses a coin and receives $5 if heads appears and $1 if tail appears. What is the expected value of a trial?
(b) A contestant receives $4.00 if a coin turns up heads and pays $3.00 if it turns tails. What is the expected value of a trail?
Answer: $\mathbf{\$3.00}$, $\mathbf{\$0.50}$</td><td></td></tr>
<tr><td>C</td><td>17</td><td>The probability distribution of a random variable X is given below. Find a, $E(X), E(2X+3), E(X^2+2), V(X), V(3X+2)$
<table>
<tr><td>X</td><td>-2</td><td>-1</td><td>0</td><td>1</td><td>2</td></tr>
<tr><td>P(x)</td><td>1/12</td><td>1/3</td><td>A</td><td>1/4</td><td>1/6</td></tr>
</table>
Answer: $\mathbf{\frac{1}{6}, \frac{1}{12}, \frac{19}{6}, \frac{43}{12}, \frac{227}{144}. \frac{227}{16}}$</td><td></td></tr>
</table>

<table>
<tr><td>H</td><td>18</td><td>The probability distribution of a random variable X is given below.
<table>
<tr><td>X</td><td>-1</td><td>0</td><td>1</td><td>2</td><td>3</td></tr>
<tr><td>P(x)</td><td>3/10</td><td>1/10</td><td>K</td><td>3/10</td><td>1/10</td></tr>
</table>
Find k, E(X), E (4X+3), $E(X^2)$, V(X), V(2X+3).
Answer: $\mathbf{1/5, 4/5, 31/5, 13/5, 49/25, 196/25}$</td><td></td></tr>
<tr><td>T</td><td>19</td><td>For a random variable X, if $E(2X-10)=20$ and $E(X^2)=400$. Find the standard deviation of X.
Answer: $\mathbf{\sqrt{175}}$</td><td></td></tr>
<tr><td>T</td><td>20</td><td>If mean and standard deviation of a random variable x are 5 and 5 respectively. Find $E(X^2)$ and $E(2X+5)^2$.
Answer: $\mathbf{E(X^2)=50, E(2X+5)^2=325}$</td><td></td></tr>
<tr><td>C</td><td>21</td><td>Find the constant c such that the function $f(x)=\begin{cases} cx^2 & ;\ 0<x<3 \\ 0 & ;\ \text{elsewhere} \end{cases}$ is a probability density function and compute $p(1<X<2)$.
Answer: $\mathbf{1/9, 7/27}$</td><td></td></tr>
<tr><td>C</td><td>22</td><td>A continuous random variable X has p.d.f. $f(x)=3x^2\ ;\ 0\le x\le 1$. Find ‘b’ such that $p(X>b)=0.05$
Answer: $\mathbf{\left(\frac{19}{20}\right)^{\frac{1}{3}}}$</td><td></td></tr>
<tr><td>C</td><td>23</td><td>A random variable X has p.d.f. $f(x)=kx^2(1-x^3)\ ;\ 0<x<1$. Find the value of ‘k’ and hence find its mean and variance.
Answer: $\mathbf{6}, 9/14, 9/245$</td><td></td></tr>
<tr><td>H</td><td>24</td><td>A random variable X has p.d.f. $f(x)=kx^2(4-x)\ ;\ 0<x<4$. Find the value of ‘k’ and hence find its mean and standard deviation.
Answer: $\mathbf{3/64, 12/5, 4/5}$</td><td></td></tr>
<tr><td>T</td><td>25</td><td>A random variable X has p.d.f. $f(x)=\begin{cases} \frac{3+2x}{18} & ;\ 2\le x\le 4 \\ 0 & ;\ \text{otherwise} \end{cases}$. Find the standard deviation of the distribution.
Answer: $\mathbf{0.57}$</td><td></td></tr>
</table>

❖ TWO DIMENSIONAL RANDOM VARIABLE

- ✓ Let S be the sample space associated with a random experiment E. Let $X = X(s)$ and $Y = Y(s)$ be two functions each assigning a real number to each outcomes. Then (X, Y) is called a two dimensional random variable.

❖ TWO DIMENSIONAL DISCRETE RANDOM VARIABLE

- ✓ If the possible values of (X, Y) are finite or countable infinite, (X, Y) is called a two dimensional discrete random variable.
  - **Example:** Consider the experiment of tossing a coin twice. The sample space S = {HH, HT, TH, TT}. Let X denotes the number of head obtained in first toss and Y denotes the number of head obtained in second toss. Then

S	HH	HT	TH	TT
X(S)	1	1	0	0
Y(S)	1	0	1	0

- ➢ Here, (X, Y) is a two-dimensional random variable and the range space of (X, Y) is $\{(1, 1), (1, 0), (0, 1), (0, 0)\}$ which is finite & so (X, Y) is a two-dimensional discrete random variables. Further,

	$Y = 0$	$Y = 1$	$Y = 2$
$X = 0$	–	–	0.25
$X = 1$	–	0.5	–
$X = 2$	0.25	–	–

❖ TWO DIMENSIONAL CONTINUOUS RANDOM VARIABLE

- ✓ If (X, Y) can assume all values in a specified region R in the xy-plane, (X, Y) is called a two dimensional continuous random variable.

❖ JOINT PROBABILITY MASS FUNCTION (DISCRETE CASE)

- ✓ If (X, Y) is a two-dimensional Discrete Random Variable such that $P(X = x_i, Y = y_j) = p_{ij}$, then p_{ij} is called the joint probability mass function of (X, Y) provided $P_{ij} \geq 0$ For all i & j and $\sum_i \sum_j P_{ij} = 1$.

❖ THE MARGINAL PROBABILITY FUNCTION (DISCRETE CASE)

- ✓ The marginal probability function is defined as

$$\mathbf{P_X(x) = \sum_y P(X = x, Y = y)}\,, \qquad \mathbf{P_Y(y) = \sum_x (X = x, Y = y)}$$

- **Example:** The joint probability mass function (PMF) of X and Y is

	$Y = 0$	$Y = 1$	$Y = 2$
$X = 0$	0.1	0.04	0.02
$X = 1$	0.08	0.2	0.06
$X = 2$	0.06	0.14	0.3

➢ The marginal Probability Mass Function of X is

$$P_X(X = 0) = 0.1 + 0.04 + 0.02 = 0.16$$
$$P_X(X = 1) = 0.08 + 0.2 + 0.06 = 0.34$$
$$P_X(X = 2) = 0.06 + 0.14 + 0.3 = 0.5$$

➢ The marginal Probability Mass Function of X is

$$P_Y(Y = 0) = 0.1 + 0.08 + 0.06 = 0.24$$
$$P_Y(Y = 1) = 0.04 + 0.2 + 0.14 = 0.38$$
$$P_Y(Y = 2) = 0.02 + 0.06 + 0.3 = 0.38$$

❖ JOINT PROBABILITY DENSITY FUNCTION (CONTINUOUS CASE)

✓ If (X,Y) is a two-dimensional continuous Random Variable such that

$$\mathbf{P\left(x - \frac{dx}{2} \le X \le x + \frac{dx}{2}, \qquad y - \frac{dy}{2} \le Y \le y + \frac{dy}{2}\right) = f(x, y)}$$

✓ It is called the joint probability density function of (X ,Y), provided

(1). $f(x, y) \ge 0$, for all $(x, y) \in D$; Where D is range of space.

(2). $\iint_D f(x, y)\, dx\, dy = 1$

✓ In particular, $P(a \le X \le b, c \le Y \le D) = \int_c^d \int_a^b f(x, y)\, dx\ dy$

❖ THE MARGINAL PROBABILITY FUNCTION (CONTINUOUS CASE)

✓ The marginal probability function is defined as

$$\mathbf{F_X(x) = \int_{-\infty}^{\infty} f(x\,, y) dy}\,, \qquad \mathbf{F_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx}$$

- **Example:** Joint probability density function of two random variables X & Y is given by

$$f(x, y) = \begin{cases} \dfrac{x^2 - xy}{8} & ; 0 < x < 2 \text{ and } -x < y < x \\ 0 & ; \text{otherw ise} \end{cases}$$

➢ The marginal probability density function of X is

$$F_X(x) = \int_{-x}^{x} f(x,y)\,dy = \int_{-x}^{x} \frac{x^2 - xy}{8}dy = \frac{1}{8}\left(x^2y - \frac{xy^2}{2}\right)_{-x}^{x} = \frac{x^3}{4}\ ; 0 < x < 2$$

- The marginal probability density function of Y is

$$F_Y(y) = \int_{0}^{2} f(x,y)\,dx = \int_{0}^{2} \frac{x^2 - xy}{8}dx = \frac{1}{8}\left(\frac{x^3}{3} - \frac{x^2y}{2}\right)_{0}^{2} = \frac{1}{3} - \frac{y}{4}\ ; -x < y < x$$

- **Remark:** The marginal distribution function of (X, Y) is

$$\mathbf{F_1(x)} = \int_{-\infty}^{x}\int_{-\infty}^{\infty} \mathbf{f(x,y)\ \ dxdy}\,, \qquad \mathbf{F_2(y)} = \int_{-\infty}^{\infty}\int_{-\infty}^{y} \mathbf{f(x,y)\ dxdy}$$

❖ INDEPENDENT RANDOM VARIABLES:

✓ Two random variables X and Y are defined to be independent if

- $P(X = x, Y = y) = P_X(x)\,.\,P_Y(y)$ If X and Y are discrete
- $f(x,y) = F_X(x)\,.\,F_Y(y)$ If X and Y are continuous
- **Example:** The joint probability mass function (PMF) of X and Y is

	$Y = 0$	$Y = 1$	$Y = 2$
$X = 0$	0.1	0.04	0.02
$X = 1$	0.08	0.2	0.06
$X = 2$	0.06	0.14	0.3

- The marginal Probability Mass Function of X=0 is

  $P(\ X\ =\ 0) = 0.1 + 0.04 + 0.02 = 0.16$

- The marginal Probability Mass Function of Y=0 is

  $P(\ Y\ =\ 0) = 0.1 + 0.08 + 0.06 = 0.24$

- $P_X(0)\ P_Y(0) = 0.16\ \times 0.24 = 0.0384$ But $P(X = 0, Y = 0) = 0.1$

  $\therefore\ \mathbf{P(X = 0, Y = 0) \neq P_X(0)\,.\,P_Y(0)}$
  ∴ X and Y are not independent random variables.

❖ EXPECTED VALUE OF TWO DIMENSIONAL RANDOM VARIABLE:

✓ Discrete case:

- $E(X) = \sum x_i\ P_X(x_i)$
- $E(Y) = \sum y_i\ P_Y(y_i)$

✓ Continuous case:

- $E(X) = \iint_R\ x\,f(x,\ y)dx\,dy$ (; where R is given region)

- $E(Y) = \iint_R \; y\, f(x,\; y) dy\, dx$ (; where R is given region)

METHOD – 6: TWO DIMENSIONAL RANDOM VARIABLE

<table>
<tr><td>C</td><td>1</td><td>The joint probability mass function of (X, Y) is given by $P(x, y) = K(2x + 3y)$ Where $x = 0,1,2$ and $y = 1,2,3$. Find the marginal probability of X.
Answer: $\mathbf{K = \frac{1}{72}}$ & $\mathbf{\frac{18}{72}, \frac{24}{72}, \frac{30}{72}}$</td><td></td></tr>
<tr><td>H</td><td>2</td><td>Let $P(X = 0, Y = 1) = 1/3, P(X = 1, Y = -1) = 1/3, P(X = 1, Y = 1) = 1/3$.
Is it the joint probability mass function of X and Y?. if yes, Find the marginal probability function of X and Y.
Answer: $\mathbf{P_X(0) = 1/3, P_X(1) = 2/3}$ & $\mathbf{P_Y(-1) = 1/3, P_Y(1) = 2/3}$</td><td></td></tr>
<tr><td>C</td><td>3</td><td>Suppose that 2 batteries are randomly chosen without replacement from the group of 12 batteries which contains 3 new batteries, 4 used batteries and 5 defective batteries. Let X denote the number of new batteries chosen and Y denote the number of used batteries chosen then find the joint probability distribution.
Answer: $\mathbf{p(0,0) = \frac{10}{66}, p(1,0) = \frac{15}{66}, p(2,0) = \frac{3}{66}, p(0,1) = \frac{20}{66}, p(1,1) = \frac{12}{66},}$
$\mathbf{p(0,2) = \frac{6}{66}}$</td><td></td></tr>
<tr><td>C</td><td>4</td><td>A two dimensional random variable (X, Y) have a bivariate distribution given by $P(X = x, Y = y) = \frac{x^2+y}{32}$ for $x = 0,1,2,3$ & $y = 0,1$. Find the marginal distributions of X and Y. Also check the independence of X & Y.
Answer: $\mathbf{X: \frac{1}{32}, \frac{3}{32}, \frac{9}{32}, \frac{19}{32},}$ $\mathbf{Y: \frac{14}{32}, \frac{18}{32},}$ NO</td><td></td></tr>
<tr><td>C</td><td>5</td><td>For given joint probability distribution of X and Y, find $p(X \le 1, Y = 2)$, $p(X \le 1), P(Y \le 3), P(X < 3, Y \le 4)$. Also check the independence of X & Y.
<table>
<tr><td></td><td>$Y = 1$</td><td>$Y = 2$</td><td>$Y = 3$</td><td>$Y = 4$</td><td>$Y = 5$</td><td>$Y = 6$</td></tr>
<tr><td>$X = 0$</td><td>0</td><td>0</td><td>1/32</td><td>2/32</td><td>2/32</td><td>3/32</td></tr>
<tr><td>$X = 1$</td><td>1/16</td><td>1/16</td><td>1/8</td><td>1/8</td><td>1/8</td><td>1/8</td></tr>
<tr><td>$X = 2$</td><td>1/32</td><td>1/32</td><td>1/64</td><td>1/64</td><td>0</td><td>2/64</td></tr>
</table>
Answer: $\mathbf{\frac{1}{16}, \frac{7}{8}, \frac{23}{64}, \frac{9}{16},}$ NO</td><td></td></tr>
</table>

<table>
<tr>
<td>H</td>
<td>6</td>
<td>The following table represents the joint probability distribution of discrete random variable (X, Y), find $P(x \leq 2, Y = 3), P(Y \leq 2)$ & $P(X + Y < 4)$.
<table>
<tr><th></th><th>$Y = 1$</th><th>$Y = 2$</th><th>$Y = 3$</th><th>$Y = 4$</th><th>$Y = 5$</th><th>$Y = 6$</th></tr>
<tr><td>$X = 0$</td><td>0</td><td>0</td><td>1/32</td><td>2/32</td><td>2/32</td><td>3/32</td></tr>
<tr><td>$X = 1$</td><td>1/16</td><td>1/16</td><td>1/8</td><td>1/8</td><td>1/8</td><td>1/8</td></tr>
<tr><td>$X = 2$</td><td>1/32</td><td>1/32</td><td>1/64</td><td>1/64</td><td>0</td><td>2/64</td></tr>
</table>
Answer: $\mathbf{\frac{11}{64}, \frac{3}{16}, \frac{13}{32}}$</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>7</td>
<td>Check the following functions are PDF or not:
(a) $f(x, y) = \begin{cases} \frac{1}{4}(1 - 6x^2y) & ; -1 \leq x < 1 \text{ and } 0 \leq y < 2 \\ 0 : & \text{otherwise} \end{cases}$
(b) $f(x, y) = \begin{cases} 3(x^2 + y^2) & ; 0 \leq x < 1 \text{ and } 0 \leq y < x \\ 0 : & \text{otherwise} \end{cases}$
Answer: YES</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>8</td>
<td>Let $f(x, y) = \begin{cases} Cxy ; & 0 < x < 4 \text{ and } 1 < y < 5 \\ 0 ; & \text{otherwise} \end{cases}$ is the joint density function of two random variables X & Y, then find the value if C.
Answer: 1/96</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>9</td>
<td>If X and Y are two random variables having joint density function:
$f(x, y) = \begin{cases} \frac{1}{8}(6 - x - y) ; & 0 \leq x < 2 \text{ and } 2 \leq y < 4 \\ 0 ; & \text{otherwise} \end{cases}$
Find $P(X < 1 \cap Y < 3), P(X + Y < 3)$.
Answer: 3/8, 5/24</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>10</td>
<td>Suppose, two dimension continuous random variable (X, Y) has Pdf given by
$f(x, y) = \begin{cases} 6x^2y ; & 0 < x < 1 \text{ and } 1 < y < 1 \\ 0 ; & \text{elsewhere} \end{cases}.$
(a) Verify: $\int_0^1 \int_0^1 f(x, y)\, dx\, dy = 1$
(b) Find $P\left(0 < X < \frac{3}{4}, \frac{1}{3} < Y < 2\right)$ & $P(X + Y < 1)$
Answer: $\mathbf{\frac{3}{8}, \frac{1}{10}}$</td>
<td></td>
</tr>
</table>

<table>
<tr><td>C</td><td>11</td><td>The joint pdf of a two-dimensional random variable (X, Y) is given by
$f(x,y)=\begin{cases}2\ ;\ \ 0<x<1,\ \ 0<y<x\\ 0\ ;\ \ \qquad\qquad\qquad \text{elsewhere}\end{cases},$
Find the marginal density function of X and Y.
Answer: $\mathbf{f_X(x)=2x, 0<x<1,\qquad f_Y(y)=2(1-y), 0<y<1}$</td><td></td></tr>
<tr><td>C</td><td>12</td><td>Check the independence of X and Y for the following:
$f(x,y)=\begin{cases}\frac{1}{4}(1+xy)\ ;\ -1<x<1,\ -1<y<1\\ 0\qquad\qquad\ ;\qquad\qquad \text{elsewhere}\end{cases}$
Answer: NO</td><td></td></tr>
<tr><td>H</td><td>13</td><td>Check the independence of X and Y for the following:
(a) $f(x,y)=\begin{cases}2\ ;\ \ 0<x<1,\ \ 0<y<x\\ 0\ ;\ \ \qquad\qquad\qquad \text{elsewhere}\end{cases}$
(b) $f(x,y)=e^{-(x+y)}\ ; 0\le x<\infty,\ \ 0\le y<\infty$
Answer: NO, YES</td><td></td></tr>
<tr><td></td><td>14</td><td>The random variables X and Y have the following joint probability distribution. What is the expected value of X and Y?
<table>
<tr><td></td><td>$Y=0$</td><td>$Y=1$</td><td>$Y=2$</td></tr>
<tr><td>$X=0$</td><td>0.2</td><td>0.1</td><td>0.2</td></tr>
<tr><td>$X=1$</td><td>0</td><td>0.2</td><td>0.1</td></tr>
<tr><td>$X=2$</td><td>0.1</td><td>0</td><td>0.1</td></tr>
</table>
Answer: $\mathbf{0.7, 1.1}$</td><td></td></tr>
<tr><td></td><td>15</td><td>Consider the joint density function for X and Y,
$f(x,y)=x^2\,y^3; 0<x<1\ \&\ 0<y<x$, find the expected value of X.
Answer: $\mathbf{1/32}$</td><td></td></tr>
</table>

UNIT-2 » SOME SPECIAL PROBABILITY DISTRIBUTIONS

❖ INTRODUCTION:

- ✓ In this chapter we shall study some of the probability distribution that figure most prominently in statistical theory and application. We shall also study their parameters. We shall introduce number of discrete probability distribution that have been successfully applied in a wide variety of decision situations. The purpose of this chapter is to show the types of situations in which these distribution can be applied.
- ✓ Some special probability distributions:
  - ➢ Binomial distribution
  - ➢ Poisson distribution
  - ➢ Normal
  - ➢ Exponential distribution
  - ➢ Gamma distribution

❖ BERNOULLI TRIALS:

- ✓ Suppose a random experiment has two possible outcomes, which are complementary, say success (S) and failure (F). If the probability $p(0 < p < 1)$ of getting success at each of the n trials of this experiment is constant, then the trials are called Bernoulli trials.

❖ BINOMIAL DISTRIBUTION:

- ✓ A random experiment consists of n Bernoulli trials such that

  (1). The trials are independent

  (2). Each trial results in only two possible outcomes, labeled as success and failure.

  (3). The probability of success in each trial remains constant.

- ✓ The random variable X that equals the number of trials that results in a success is a binomial random variable with parameters $0 < p < 1, q = 1 - p$ and $n = 1, 2, 3, \ldots.$ The probability mass function of X is

$$P(X = x) = \binom{n}{x} p^x \, q^{n-x} \qquad ; x = 0, 1, 2, \ldots \ldots, n.$$

❖ EXAMPLES OF BINOMIAL DISTRIBUTION:

(1). Number of defective bolts in a box containing n bolts.

(2). Number of post-graduates in a group of n people.

(3). Number of oil wells yielding natural gas in a group of n wells test drilled.

(4). In the next 20 births at a hospital. Let X=the number of female births.

(5). Flip a coin 10 times. Let X=number of heads obtained.

- **NOTE:**
  - The mean of binomial distribution is defined as $\boldsymbol{\mu = E(X) = np.}$
  - The variance of the binomial distribution is defined as $\mathbf{V(X) = npq}$ & $\mathbf{S.D.}\,\boldsymbol{\sigma} = \sqrt{\mathbf{npq}}$
  - $\binom{n}{x} = \text{combinaton} = \frac{n!}{(n-x)!\,x!}$
  - $\binom{n}{x} = \binom{n}{n-x}$
  - $\binom{n}{0} = \binom{n}{n} = 1$

METHOD – 1: BINOMIAL DISTRIBUTION

H	1	Write assumption of Binomial Distribution.	
C	2	Find the binomial distribution for $n = 4$ and $p = 0.3$. **Answer**: $\mathbf{0.2401, 0.4116, 0.2646, 0.0756, 0.0081}$	
C	3	12% of the tablets produced by a tablet machine are defective. What is the probability that out of a random sample of 20 tablets produced by the machine, 5 are defective? **Answer:** $\mathbf{0.0567}$	
H	4	20% Of the bulbs produced are defective .Find probability that at most 2 bulbs out of 4 bulbs are defective. **Answer:** $\mathbf{0.9728}$	
C	5	If 3 of 12 car drivers do not carry driving license, what is the probability that a traffic inspector who randomly checks 3 car drivers, will catch 1 for not carrying driving license. (Use binomial dist.) **Answer**: $\mathbf{\frac{27}{64}}$	

H	6	The probability that India wins a cricket test match against Australia is given to be $\frac{1}{3}$. If India and Australia play 3 tests matches, what is the probability that (a) India will lose all the three test matches. (b) India will win at least one test match. **Answer:** (a) $\mathbf{0.2963}$, (b) $\mathbf{0.7037}$	
T	7	What are the properties of Binomial Distribution? The average percentage of failure in a certain examination is 40. What is the probability that out of a group of 6 candidates, at least 4 passed in examination? **Answer:** $\mathbf{0.5443}$	
C	8	The probability that in a university, a student will be a post-graduate is 0.6. Determine the probability that out of 8 students (a) None (b) Two (c) At least two will be post-graduate **Answer:** (a) $\mathbf{0.0007}$, (b) $\mathbf{0.0413}$, (c) $\mathbf{0.9914}$	
H	9	The probability that an infection is cured by a particular antibiotic drug within 5 days is 0.75. Suppose 4 patients are treated by this antibiotic drug. What is the probability that (a) no patient is cured (b) exactly two patient are cured (c) At least two patients are cured. **Answer:** (a) $\mathbf{0.0039}$, (b)$\mathbf{0.2109}$, (c)$\mathbf{0.9492}$	
C	10	Assume that on the average one telephone number out of fifteen called between 1 p.m. and 2 p.m. on weekdays is busy. What is the probability that if 6 randomly selected telephone numbers were called (a) not more than three, (b) at least three of them would be busy? **Answer:** $\mathbf{0.9997, 0.0051}$	
H	11	A dice is thrown 6 times getting an odd number of success, Find probability (a) Five success (b) At least five success (c) At most five success. **Answer:** (a) $\mathbf{\frac{3}{32}}$, (b) $\mathbf{\frac{7}{64}}$, (c) $\mathbf{\frac{63}{64}}$	

T	12	A multiple choice test consist of 8 questions with 3 answer to each question (of which only one is correct). A student answers each question by rolling a balanced dice & checking the first answer if he gets 1 or 2, the second answer if he gets 3 or 4 & the third answer if he gets 5 or 6. To get a distinction, the student must secure at least 75% correct answers. If there is no negative marking, what is the probability that the student secures a distinction? **Answer**: $\mathbf{P(X \geq 6) = 0.0197}$	
H	13	Obtain the binomial distribution for which mean is 10 and variance is 5. **Answer**: $\mathbf{P(X = x) = \binom{20}{x}(0.5)^{x}(0.5)^{20-x}}$	
C	14	Determine binomial distribution whose mean is 4 and variance is 3 and hence evaluate $P(X \geq 2)$. **Answer**: $\mathbf{P(X = x) = \begin{cases} \binom{16}{x}(0.25)^{x}(0.75)^{16-x} & ; x = 0, 1, \dots, 16 \\ 0 & ; otherwise \end{cases}}$ $\mathbf{P(X \geq 2) = 0.9365}$	
H	15	For the binomial distribution with $n = 20, p = 0.35$. Find Mean, Variance and Standard deviation. **Answer**: $\mathbf{(a)7.0000\ (b)4.55, (c)2.1331}$	
C	16	If the probability of a defective bolt is 0.1 Find mean and standard deviation of the distribution of defective bolts in a total of 400. **Answer**: $\mathbf{\mu = 40, \sigma = 6}$	

❖ POISSON DISTRIBUTION:

✓ A discrete random variable X is said to follow Poisson distribution if it assume only non-negative values and its probability mass function is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!};\ x = 0, 1, 2, 3, \dots . \text{and } \lambda = \text{mean of the Poisson distribution}$$

❖ EXAMPLES OF POISSON DISTRIBUTION:

(1). Number of defective bulbs produced by a reputed company.

(2). Number of telephone calls per minute at a switchboard.

(3). Number of cars passing a certain point in one minute.

(4). Number of printing mistakes per page in a large text.

(5). Number of persons born blind per year in a large city.

❖ PROPERTIES OF POISSON DISTRIBUTION:

✓ The Poisson distribution holds under the following conditions.

- The random variable X should be discrete.
- The number of trials n is very large.
- The probability of success p is very small (very close to zero).
- The occurrences are rare.
- $\lambda = \text{np}$ is finite.
- The mean and variance of the Poisson distribution with parameter λ are defined as follows. $\mathbf{Mean}\ \boldsymbol{\mu} = \mathbf{E(X)} = \boldsymbol{\lambda} = \mathbf{n\,p}$; $\mathbf{Variance}\ \mathbf{V(X)} = \boldsymbol{\sigma^2} = \boldsymbol{\lambda}.$

METHOD – 2: POISSON DISTRIBUTION

H	1	In a company, there are 250 workers. The probability of a worker remain absent on any one day is 0.02. Find the probability that on a day seven workers are absent. **Answer**: $\mathbf{P(X = 7) = 0.104}$	
C	2	A book contains 100 misprints distributed randomly throughout its 100 pages. What is the probability that a page observed at random contains at least two misprints. Assume Poisson Distribution. **Answer**: $\mathbf{0.2642}$	
C	3	For Poisson variant X, if $P(X = 3) = P(X = 4)$ then, Find $P(X = 0)$. **Answer**: $\mathbf{P(X = 0) = e^{-4}}$	
C	4	In a bolt manufacturing company, it is found that there is a small chance of $\frac{1}{500}$ for any bolt to be defective. The bolts are supplied in a packed of 30 bolts. Use Poisson distribution to find approximate number of packs, (a) Containing no defective bolt and (b) Containing two defective bolt, in the consignment of 10000 packets. **Answer**: (a) $\mathbf{0.9802}$; (b) $\mathbf{17}$	

H	**5**	In sampling a large number of parts manufactured by a machine, the mean number and of defectives in a sample of 20 is 2. Out of 1000 such samples, how many would be expected to contain exactly two defective parts? **Answer**: $\mathbf{P(X = 2) = 0.270}$	
C	**6**	Potholes on a highway can be serious problems. The past experience suggests that there are, on the average, 2 potholes per mile after a certain amount of usage. It is assumed that the Poisson process applies to the random variable “no. of potholes”. What is the probability that no more than four potholes will occur in a given section of 5 miles? **Answer**: $\mathbf{P(X \leq 4) = 0.0315}$	
H	**7**	A car hire firm has two cars, which it hires out day by day. The number of demands for a car on each day is distributed on a Poisson distribution with mean 1.5. Calculate the proportion of days on which neither car is used and proportion of days on which some demand is refused. $(e^{-1.5} = 0.2231)$. **Answer**: $\mathbf{P(X = 0) = 0.2231\ ;\ 1 - P(X \leq 2) = 0.1912}$	
H	**8**	100 Electric bulbs are found to be defective in a lot of 5000 bulbs. Find the probability that at the most 3 bulbs are defective in a box of 100 bulbs. **Answer**: $\mathbf{P(X \leq 3) = 0.8567}$	
C	**9**	Certain mass produced articles of which 0.5% are defective ,are packed in cartons each containing 100. What proportion of cartons are free from defective articles, and what proportion contain 3 or more defectives? **Answer**: $\mathbf{P(X = 0) = 60.65, P(X \geq 3) = 0.0144}$	
H	**10**	If a bank receives an average six back cheques per day what are the probability that bank will receive (a) 4 back cheques on any given day (b) 10 back cheques on any consecutive day. **Answer**: $\mathbf{(a)\ P(X = 4) = 0.1338\ (b)\ 2P(X = 10) = 0.0826}$	
C	**11**	The probability that a person catch swine flu virus is 0.001. Find the probability that out of 3000 persons (a) exactly 3, (b) more than 2 person will catch the virus. (F(2; λ)=0.42) **Answer: (a) 0.2242 (b) 0.5769**	
T	**12**	Suppose 1% of the items made by machine are defective. In a sample of 100 items find the probability that the sample contains All good, 1 defective and at least 3 defective. **Answer**: $\mathbf{P(X = 0) = 0.3679;\ P(X = 1) = 0.3679;\ P(X \geq 3) = 0.0803}$	

❖ EXPONENTIAL DISTRIBUTION:

- ✓ A random variable X is said to have an Exponential distribution with parameter θ > 0, if its probability density function is given by:

$$f(X=x)=\begin{cases}\theta\, e^{-\theta x}\ ;\ x\geq 0\\ 0\ ;\ otherwise\end{cases}$$

- ✓ Here, $\theta=\frac{1}{\text{mean}}$ or mean $=\frac{1}{\theta}$ and variance $=\frac{1}{\theta^2}$.
- ✓ Exponential distribution is a special case of Gamma distribution.
- ✓ Exponential distribution is used to describe lifespan and waiting times.
- ✓ Exponential distribution can be used to describe (waiting) times between Poisson events.

METHOD – 3: EXPONENTIAL DISTRIBUTION

C	1	The lifetime T of an alkaline battery is exponentially distributed with θ = 0.05 per hour. What are mean and standard deviation of batteries lifetime? **Answer: 20 hr**	
C	2	The lifetime T of an alkaline battery is exponentially distributed with θ = 0.05 per hour. What are the probabilities for battery to last between 10 and 15 hours? What are the probabilities for the battery to last more than 20 hr? **Answer: 0.1341, 0.3679**	
H	3	The time between breakdowns of a particular machine follows an exponential distribution with a mean of 17 days. Calculate the probability that a machine breakdown in 15 days period. **Answer: 0.5862**	
H	4	The arrival rate of cars at a gas station is 40 customers per hour. (a) What is the probability of having no arrivals in 5 min. interval? (b) What is the probability for having 3 arrivals in 5 min.? **Answer: (a) 0.03567, (b) 0.2202**	

C	**5**	In a large corporate computer network, user log-on to the system can be modeled as a Poisson process with a mean of 25 log-on per hours. (a) What is the probability that there are no log-on in an interval of six min.? (b) What is the probability that time until next log-on is between 2 & 3 min.? (c) Determine the interval of time such that the probability that no log-on occurs in the interval is 0.90? **Answer: (a) 0.07788, (b) 0.152, (c) 0.25**	
H	**6**	The time intervals between successive barges passing a certain point on a busy waterway have an exponential distribution with mean 8 minutes. (a) Find the probability that the time interval between two successive barges is less than 5 minutes. (b) Find a time interval t such that we can be 95% sure that the time interval between two successive barges will be greater than t. **Answer: 0.4647, 24.6sec**	
T	**7**	Accidents occur with Poisson distribution at an average of 4/week ($\lambda = 4$). (a) Calculate the probability of more than 5 accidents in any one week. (b) What is probability that at least two weeks will elapse between accidents? **Answer: (a) 0.215, (b) 0.00034**	

❖ GAMMA DISTRIBUTION:

- ✓ A random variable X is said to have a Gamma distribution with parameter $n > 0$, if its probability density function is given by:

$$f(X = x) = \begin{cases} \dfrac{\lambda^r\, x^{r-1}\, e^{-\lambda x}}{\Gamma(n)} & ;\ 0 < x < \infty \ and\ r = 1, 2, \dots \\ 0 & ;\qquad otherwise \end{cases}$$

- ✓ Here, mean $= \frac{r}{\lambda}$ and variance $= \frac{r}{\lambda^2}$.

METHOD – 4: GAMMA DISTRIBUTION

H	**1**	Given a gamma random variable X with $r = 3$ and $\lambda = 2$. Compute E(X), V(X) and $P(X \leq 1.5$ years). **Answer: 1.5, 0.75, 0.5768**	

C	**2**	Suppose you are fishing and you expect to get a fish once every 1/2 hour. Compute the probability that you will have to wait between 2 to 4 hours before you catch 4 fish. **Answer: 0.124**	
C	**3**	The daily consumption of milk in a city, in excess of 20000 liters, is approximately distributed as a gamma variate with r=2 and λ=1/10000. The city has daily stock of 30000 liters. What is the probability that the stock is insufficient on a particular day? **Answer: 0.736**	
T	**4**	The daily consumption of electric power in a certain city is a random variable X having probability density function $f(x) = \begin{cases} \frac{1}{9}\, x\, e^{-\frac{x}{3}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$. Find the probability that the power supply is inadequate on any given day if the daily capacity of the power plant is 12 million KW hours. **Answer: 0.09758**	
H	**5**	Suppose that the time it takes to get service in a restaurant follows a gamma distribution with mean 8 min and standard deviation 4 minutes. (a) Find the parameters r and λ of the gamma distribution. (b) You went to this restaurant at 6:30. What is the probability that you will receive service before 6:36? **Answer: 2, 1/4, 0.442**	

❖ NORMAL DISTRIBUTION:

- ✓ A continuous random variable X is said to follow a normal distribution of its probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]; -\infty < x < \infty, \sigma > 0$$

- ✓ Where, μ = mean of the distribution and σ = Standard deviation of the distribution
  - **Note:** μ and σ^2 (variance) are called parameters of the distribution.
- ✓ If X is a normal random variable with mean $\boldsymbol{\mu}$ and standard deviation $\boldsymbol{\sigma}$, and if we find the random variable $\mathbf{Z} = \frac{X-\mu}{\sigma}$ with mean 0 and standard deviation 1, then Z in called the standard (standardized) normal variable.

- ✓ The probability destiny function for the normal distribution in standard form is given by

$$\boldsymbol{f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}; -\infty<z<\infty}$$

- ✓ The distribution of any normal variate X can always be transformed into the distribution of the standard normal variate Z.

$$P(x_1 \le X \le x_2)=P\left(\frac{x_1-\mu}{\sigma}\le\frac{X-\mu}{\sigma}\le\frac{x_2-\mu}{\sigma}\right)=P(z_1\le Z\le z_2)$$

- ✓ For normal distribution

  (1) $\mathrm{P}(-\infty \le \mathrm{z} \le \infty)=1$ (Total area)

  (2) $\mathrm{P}(-\infty \le \mathrm{z} \le 0)=\mathrm{P}\ (0 \le \mathrm{z} \le \infty)=0.5$

  (3) $\mathrm{P}(-\mathrm{z}_1 \le \mathrm{z} \le 0)=\mathrm{P}(0 \le \mathrm{z} \le \mathrm{z}_1)\ \ ;\ \mathrm{z}_1>0$

METHOD – 5: NORMAL DISTRIBUTION

<table>
<tr>
<td>C</td>
<td>1</td>
<td>Compute the value of following:
<table>
<tr><td>$p(0 \le z \le 1.43)$</td><td>$p(-0.73 \le z \le 0)$</td></tr>
<tr><td>$p(-1.37 \le z \le 2.02)$</td><td>$p(0.65 \le z \le 1.26)$</td></tr>
<tr><td>$p(z \ge 1.33)$</td><td>$p(|z| \le 0.5)$</td></tr>
</table>
Answer: 0.4236, 0.2673, 0.8930, 0.154, 0.0918, 0.383</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>2</td>
<td>What is the probability that a standard normal variate Z will be greater than 1.09, less than -1.65, lying between -1 & 1.96, lying between 1.25 & 2.75?
Answer: 0.1379, 0.0495, 0.8163, 0.1026</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>3</td>
<td>The compressive strength of the sample of cement can be modelled by normal distribution with mean 6000 $\mathrm{kg/cm^2}$ and standard deviation of 100 $\mathrm{kg/cm^2}$.
(a) What is the probability that a sample strength is less than 6250 $\mathrm{kg/cm^2}$?
(b) What is probability if sample strength is between 5800 and 5900 $\mathrm{kg/cm^2}$?
(c) What strength is exceeded by 95% of the samples?
$[\mathrm{P(z=2.5)=0.9938, P(z=1)=0.8413, P(z=2)=0.9772, P(z=1.65)=0.95}]$
Answer: (a)0.9938 , (b) 0.1815 , (c) 1.65 , x = 6165</td>
<td></td>
</tr>
</table>

C	4	In a photographic process, the developing time of prints may be looked upon as a random variable having normal distribution with mean of 16.28 seconds and standard deviation of 0.12 seconds. Find the probability that it will take **(a)** Anywhere from 16.00 to 16.50 sec to develop one of the prints; **(b)** At least 16.20 sec to develop one of the prints; **(c)** At most 16.35 sec to develop one of the prints. $[P(z = 1.83) = 0.9664, P(z = 0.66) = 0.7454, P(z = 0.58) = 0.7190]$ **Answer**: (a) $\mathbf{0.9565}$, (b) $\mathbf{0.7475}$, (c) $\mathbf{0.7190}$	
C	5	A random variable having the normal distribution with $\mu = 18.2$ & $\sigma = 1.25$, find the probabilities that it will take on a value **(a)** less than 16.5 **(b)** Between 16.5 and 18.8. [F(0.48)=0.3156, F(-1.36)=0.0869] **Answer**: **(a)** $\mathbf{0.0885}$ **(b)** $\mathbf{0.5959}$	
H	6	A sample of 100 dry battery cell tested and found that average life of 12 hours and standard deviation 3 hours. Assuming the data to be normally distributed what percentage of battery cells are expected to have life **(a)** More than 15 hour **(b)** Less than 6 hour **(c)** Between 10 & 14 hours **Answer**: (**a**) $\mathbf{15.87\%}$, (**b**) $\mathbf{2.28\%}$, (**c**) $\mathbf{49.72\%}$	
C	7	In AEC company, the amount of light bills follows normal distribution with standard deviation 60. 11.31% of customers pay light-bill less than Rs.260. Find average amount of light bill. **Answer**: $\boldsymbol{\mu = 332.60}$	
H	8	The marks obtained by students in a college are normally distributed with mean of 65 & variance of 25. If 3 students are selected at random from the college, what is the probability that at least one of them would have scored more than 75 marks? **Answer**: $\mathbf{0.0668}$	
C	9	Weights of 500 students of college is normally distributed with average weight 95 lbs. & σ =7.5. Find how many students will have the weight between 100 and 110. **Answer**: $\mathbf{114}$	

H	**10**	Distribution of height of 1000 soldiers is normal with Mean 165cm & standard deviation 15cms how many soldiers are of height **(a)** Less than 138 cm, **(b)** More than 198 cm, **(c)** Between 138 & 198 cm **Answer**: (a) **36**, (b) **14**, (c) **950**	
C	**11**	In a normal distribution 31% of the items are below 45 and 8% are above 64. Determine the mean and standard deviation of this distribution **Answer**: $\mathbf{\mu = 49.974, \sigma = 9.95}$	
H	**12**	The breaking strength of cotton fabric is normally distributed with $E(x) = 16$ and $\sigma(x) = 1$. The fabric is said to be good if $x \geq 14$ what is the probability that a fabric chosen at random is good? **Answer**: $\mathbf{0.9772}$	

❖ BOUNDS ON PROBABILITIES

- ✓ If the probability distribution of a random variable is known $E(X)$ & $V(x)$ can be computed. Conversely, if $E(X)$ & $V(X)$ are known, probability distribution of X cannot be constructed and quantities such as $P\{|X - E(X)| \leq K\}$ cannot be evaluate.
- ✓ Several approximation techniques have been developed to yield upper and/or lower bounds to such probabilities. The most important of such technique is Chebyshev's inequality.

❖ CHEBYSHEV'S INEQUALITY

- ✓ If X is a random variable with mean μ and variance σ^2, then for any positive number k,

$$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2} \quad or \quad P\{|X-\mu| < k\sigma\} \geq 1 - \frac{1}{k^2}$$

METHOD – 6: CHEBYSHEV'S INEQUALITY

C	**1**	A random variable X has a mean 12, variance 9 and unknown probability distribution. Find $P(6 < X < 18)$. **Answer:** $\mathbf{P(6 < X < 18) \geq \frac{3}{4}}$	

H	**2**	The number of customers who visit a car dealer's showroom on Sunday morning is a random variable with mean 18 & $\sigma = 2.5$. What is the probability that on Sunday morning the customers will be between 8 to 28? **Answer:** $\mathbf{P(8 < X < 28) \geq \frac{15}{16}}$	
C	**3**	Determine the smallest value of 'k' in the chebyshev's inequality for which the probability is at least 0.95. **Answer:** $\mathbf{k = \sqrt{20}}$	
H	**4**	A random variable X has mean 10, variance 4 and unknown probability distribution. Find 'c' such that $P\{\lvert X - 10\rvert \geq c\} \leq 0.04$. **Answer: 10**	
T	**5**	A random variable X has pdf $f(x) = e^{-x}, x \geq 0$. Use chebyshev's inequality to show that $P\{\lvert X - 1\rvert > 2\} \leq \frac{1}{4}$ and also find the actual probability. **Answer:** $\mathbf{1 - e^{-3}}$	

❖ STANDARD NORMAL (Z) TABLE , AREA BETWEEN 0 AND Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3990	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4115	0.4131	0.4147	0.4162
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

UNIT-3 » BASIC STATISTICS

❖ INTRODUCTION:

- ✓ Statistics is the branch of science where we plan, gather and analyze information about a particular collection of object under investigation. Statistics techniques are used in every other field of science, engineering and humanity, ranging from computer science to industrial engineering to sociology and psychology.
- ✓ For any statistical problem the initial information collection from the sample may look messy, and hence confusing. This initial information needs to be organized first before we make any sense out of it.
- ✓ Quantitative data in a mass exhibit certain general characteristics or they differ from each other in the following ways:
  - ➢ They show a tendency to concentrate values, usually somewhere in the centre of the distribution. Measures of this tendency are called measures of **Central Tendency**.
  - ➢ The data vary about a measure of Central tendency and these measures of deviation are called measures of variation or **Dispersion.**
  - ➢ The data in a frequency distribution may fall into symmetrical or asymmetrical patterns. The measure of the direction and degree of asymmetry are called measures of **Skewness**.
  - ➢ Polygons of frequency distribution exhibit flatness or peakedness of the frequency curves. The measures of peakedness of the frequency curves are called measures of **Kurtosis**.

❖ DATA

- ✓ Data is a set of values of qualitative or quantitative variables.
- ✓ Types of data (On the basis of Numbers):
  - ➢ **Univariate Data:** If one variable is required for observation, than it is called Univariate data. e.g. $\{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \dots, \mathbf{x_n}\}$.
  - ➢ **Bivariate Data:** If two variables is required for observation, than it is called Univariate data. e.g. $\{(\mathbf{x_1}, \mathbf{y_1}), (\mathbf{x_2}, \mathbf{y_2}), (\mathbf{x_3}, \mathbf{y_3}), \dots, (\mathbf{x_n}, \mathbf{y_n})\}$.
  - ➢ **Multivariate Data:** If more than one variables is required to describe the data, than it is called Multivariate data.

UNIVARIATE ANALYSIS:

- ✓ Univariate analysis involves the examination across cases of one variable at a time. There are three major characteristics of a single variable that we tend to look at:
  - Distribution (Data)
  - Central Tendency
  - Dispersion

DISTRIBUTION:

- ✓ Distribution of a statistical data set (or a population) is a listing or function showing all the possible values (or intervals) of the data and how often they occur.
- ✓ Type of distribution (Data):
  - Distribution of ungrouped data:

    E.g. Marks of AEM of **10** students are $\mathbf{10, 25, 26, 35, 03, 08, 19, 29, 30, 18}$.
  - Distribution of grouped data:
    - Discrete Frequency Distribution:

      E.g. Data of Students using Library during exam time.

No. reading hours(x_i)	1	2	3	4
No. of hostel students(f_i)	4	7	10	8

- Continuous Frequency Distribution:

  E.g. Data of Students using Library during exam time.

No. reading hours (x_i)	$0 - 2$	$3 - 5$	$6 - 8$	$9 - 11$
No. of hostel students (f_i)	11	7	8	0

SOME DEFINATION:

- ✓ **Exclusive Class:** If classes of frequency distributions are $\mathbf{0-2, 2-4, 4-6}, \dots$ such classes are called Exclusive Classes.
- ✓ **Inclusive Class:** If classes of frequency distributions are $\mathbf{0-2, 3-5, 6-8}, \dots$ such classes are called Inclusive Classes.
- ✓ **Lower Boundary & Upper Boundary:** In Class $\mathbf{x_i - x_{i+1}}$, Lower Boundary is $\mathbf{x_i}$ and Upper boundary is $\mathbf{x_{i+1}}$.
- ✓ **Mid-Point of class:** It is defined as

$$\frac{\textbf{Lower Boundary} + \textbf{Upper Boundary}}{2}$$

❖ CENTRAL TENDENCY:

- ✓ The central tendency of a distribution is an estimate of the "center" of a distribution of values. There are three major types of estimates of central tendency:
  - ➢ Mean
  - ➢ Median
  - ➢ Mode

❖ MEAN($\bar{x}$):

- ✓ The Mean or Average is probably the most commonly used method of describing central tendency. To compute the mean, add up all the values and divide by the number of values.
- ✓ Mean for Ungrouped data:
  - ➢ If data is $x_1, x_2, x_3, \dots, x_n$. Then Mean $\bar{x}$ is defined as below.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

- ✓ Mean for Grouped data:
  - ➢ Direct Method:
    - Suppose, given data (OR Mid Values of Class) is $x_1, x_2, x_3, \dots, x_n$. Mean is defined as below.

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \cdots + x_n f_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i f_i$$

  - ➢ Assumed Mean Method:
    - Suppose, given data (OR Mid Values of Class) is $x_1, x_2, x_3, \dots, x_n$. Let $A = x_i$. Mean is defined as below.

$$\bar{x} = A + \frac{1}{n} \cdot \sum_{i=1}^{n} d_i f_i \quad ; \quad \textbf{Where } d_i = x_i - A \textbf{ and A can be any value of } x_i$$

  - ➢ Step Deviation Method:
    - Suppose, given data (OR Mid Values of Class) is $x_1, x_2, x_3, \dots, x_n$ with class length **C**. Let $A = x_i$. Mean is defined as below.

$$\bar{x} = A + \frac{c}{n} \cdot \sum_{i=1}^{n} u_i f_i \quad ; \quad \textbf{Where } u_i = \frac{x_i - A}{c}$$

- Step deviation method is used for only continuous frequency distribution.

❖ MODE (Z):

- ✓ The Mode is the most frequently occurring value in the set. To determine the mode, you might again order the observations in numerical order and then count each one. The most frequently occurring value is the mode.
- ✓ Mode of Ungrouped data:
    - ➢ Most repeated observation among given data is called Mode of Ungrouped data.
- ✓ Mode of Grouped data:
    - ➢ Discrete Frequency Distribution:
        - The value of variable corresponding to maximum frequency.
    - ➢ Continuous Frequency Distribution:

$$\mathbf{Z = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times c}$$

l = Lower boundary of Modal Class

c = class interval OR class length

f_1 = Frequency of the modal class.

f_0 = Frequency of the class preceding the modal class.

f_2 = Frequency of the class succeeding the modal class.

Where, the modal class is the class with highest frequency

❖ MEDIAN (M):

- ✓ The Median is the value found at the exact middle of the set of values. To compute the median is to list all observations in numerical order and then locate the value in the center of the sample.
- ✓ Median of Ungrouped Data:
    - ➢ Let the total number of observation be **n**.
    - ➢ $\mathbf{If\ n\ is\ odd\ number, than\ M = \left(\frac{n+1}{2}\right)^{th}\ observation.}$
    - ➢ $\mathbf{If\ n\ is\ even\ number, than\ M = \frac{\left(\frac{n}{2}\right)^{th} observation + \left(\frac{n}{2}+1\right)^{th} observation}{2}}.$
- ✓ Median of Grouped Data:
    - ➢ Median of Discrete Grouped Data:

- In case of discrete group data the position of median i.e. $\left(\frac{n+1}{2}\right)^{th}$ item can be located through cumulative frequency. The corresponding value at this position is value of median.

➢ Median Of Continuous Grouped Data:

$$\mathbf{M} = \mathbf{l} + \left(\frac{\frac{\mathbf{n}}{\mathbf{2}} - \mathbf{F}}{\mathbf{f}}\right) \times \mathbf{c}$$

$l =$ lower boundary point of the Median class

$n =$ total number of observation (sum of the frequencies)

$F =$ cumulative frequency of the class preceding the median class.

$f =$ the frequency of the median class

$c =$ class interval **OR** class length

Where, Median class = Class whose cumulative frequency with property $\mathbf{min}\left\{\mathbf{cf} \mid \mathbf{cf} \geq \frac{\mathbf{n}}{\mathbf{2}}\right\}$

❖ DISPERSION:

✓ Dispersion refers to the spread of the values around the central tendency. There are two common measures of dispersion, the range and the standard deviation.

✓ **Range:** It is simply the highest value minus the lowest value. In our example distribution, the high value is **36** and the low is **15**, so the range is $\mathbf{36 - 15 = 21}$.

✓ **Standard Deviation($\boldsymbol{\sigma}$):** It is a measure that is used to quantify the amount of variation or dispersion of a set of data values.

❖ FORMULA TO FIND STANDARD DEVIATION:

Method	Ungrouped Data	Grouped Data
Direct Method	$\sigma = \sqrt{\frac{\sum_{j=1}^{N}\left(x_j - \bar{x}\right)^2}{n}}$	$\sigma = \sqrt{\frac{\sum_{j=1}^{N} f_j\left(x_j - \bar{x}\right)^2}{n}}$
Actual Mean Method	$\sigma = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2}$	$\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i x_i^2}{\sum_{i=1}^{n} f_i} - \left(\frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}\right)^2}$
Assumed Mean Method	$\sigma = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n} - \left(\frac{\sum_{i=1}^{n} d_i}{n}\right)^2}$	$\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i d_i^2}{\sum_{i=1}^{n} f_i} - \left(\frac{\sum_{i=1}^{n} f_i d_i}{\sum_{i=1}^{n} f_i}\right)^2}$

❖ VARIANCE (V(x)):

✓ Variance is expectation of the squared deviation. It informally measures how far a set of (random) numbers are spread out from their mean.

$$\textbf{Variance} = (\textbf{Standard Deviation})^2$$

❖ SKEWNESS:

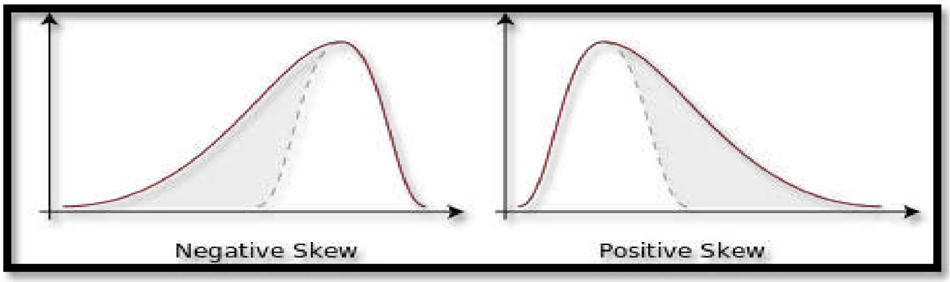
- ✓ It is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean.

$$\textbf{Skewness} = \frac{\textbf{Mean} - \textbf{Mode}}{\textbf{Standard deviation}} = \frac{\bar{x} - Z}{\sigma}$$

- ✓ The Skewness value can be positive or negative, or even undefined.
- ✓ Negative skew:

  The left tail is longer; the mass of the distribution is concentrated on the right.
- ✓ Positive skew:

  The right tail is longer; the mass of the distribution is concentrated on the left.



❖ COEFFICIENT OF VARIATION

- ✓ Formula: $\text{C.V.} = \frac{\sigma}{\bar{x}} \times \mathbf{100}$
- ✓ The Coefficient of variation is lesser is said to be less variable or more consistent.

METHOD – 1: CENTRAL TENDENCY, DISPERSION AND SKEWNESS

C	1	Define Mode and also give the relationship between Mean, Median & Mode. **Answer**: $\mathbf{Z = 3M - 2\bar{X}}$	
C	2	Find the mean, mode and median by the data 2, 8, 4, 6, 10, 12, 4, 8, 14, 16. **Answer**: $\mathbf{\bar{X} = 8.4, Z = 8}$ **&** $\mathbf{4, M = 8}$	
H	3	Find mean, mode and median of following observation. 10, 9, 21, 16, 14, 18, 20, 18, 14, 18, 23, 16, 18,4 . **Answer**: $\mathbf{15.6429, 18, 17}$	

C **4** Find mean, mode and Median of following data.

Number of students	6	4	16	7	8	2
Marks obtained	20	9	25	50	40	80

Answer: $\mathbf{32.2326, 25, 25}$

H **5** Find mean, mode and Median of following data

Marks obtained	18	22	30	35	39	42	45	47
Number of students	4	5	8	8	16	4	2	3

Answer: $\mathbf{34.5, 39, 37}$

C **6** The following data represents the no. of foreign visitors in a multinational company in every 10 days during last 2 months. Use the data to the mean, mode and median.

X	$0-10$	$10-20$	$20-30$	$30-40$	$40-50$	$50-60$
No. of visitors	12	18	27	20	17	06

Answer: $\mathbf{28, 25.625, 30.7407}$

H **7** A Survey regarding the weights (in kg) of 45 students of class X of a school was conducted and the following data was obtained:

Weight (in kg)	20-25	25-30	30-35	35-40	40-45	45-50	50-55
No. of students	2	5	8	10	7	10	3

Find the mean, mode and median weight.

Answer: $\mathbf{38.8333, 37, 38.75}$

H **8** Find the missing frequency when median is 24.

Marks	$0-10$	$10-20$	$20-30$	$30-40$	$40-50$
Students	15	20	X	14	16

Answer: $\mathbf{25}$

H **9** Obtain the mean, mode and median for the following distribution:

Mid value	15	20	25	30	35	40	45	50	55
Frequency	2	22	19	14	3	4	6	1	1
Cumulative	2	24	43	57	60	64	70	71	72

Answer: 25.8472, 21.8478, 25.6579

H **10** Obtain the mean, mode and median for the following information:

Marks	0 <	10 <	20 <	30 <
Students	50	38	20	5

Answer: 17.6, 16.6667, 17.2222

T **11** Obtain the mean, mode and median for the following information:

Daily wage	< 50	50 – 100	100 – 150	150 – 200	200 – 250	250 – 300	300 – 350
Frequency	2	4	7	21	25	20	21

Answer: 228.5, 222.2222, 232

T **12** Calculate Mean, Median and Mode for the following data.

Class	50 – 53	53 – 56	56 – 59	59 – 62	62 – 65	65 – 68	68 – 71	71 – 74	74 – 77
F	3	8	14	30	36	28	16	10	5

Answer: 63.82, 63.66, 63.2857

C **13** The pH of a solution is measured 7 times by one operator using the same instrument. She obtains the following data: 7.15, 7.20, 7.18, 7.19, 7.21, 7.16 and 7.18. Calculate Skewness.

Answer: 0.0496

H **14** Find Skewness of temperature recorded in degree centigrade during a week in May, 2015, where the temperature recorded are 38.2, 40.9, 39.5, 44, 39.6, 40.5, 39.5.

Answer: 0.4261

C **15** Calculate the Skewness:

X	20	9	25	50	40	80
f	6	4	16	7	8	2

Answer: 0.4492

H **16** Find Skewness from the following table.

X	35	45	55	60	75	80
f	12	18	10	6	3	11

Answer: 0.5788

C **17** Find Skewness from the following data.

Class	9 – 11	12 – 14	15 – 17	18 – 20
Frequency	2	3	4	1

Answer: 2.74955

H **18** Find the Skewness of the data given below by all the three methods:

Class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
Frequency	4	8	3	20	3	4	8

Answer: 0.044493

H **19** Find the Standard deviation and variance of the mark distribution of 30 students at mathematics examination in a class as below:

Marks	10-25	25-40	40-55	55-70	70-85	85-100
No of students	05	21	21	08	25	20

Answer: 23.8736, 569.9488

H **20** Runs scored by two batsman A, B in 9 consecutive matches are given below:

A	85	20	62	28	74	5	69	4	13
B	72	4	15	30	59	15	49	27	26

Which of the batsman is more consistent?

Answer: Batsman B is more consistent

T **21** Two machines A, B are used to fill a mixture of cement concrete in a beam.

A	32	28	47	63	71	39	10	60	96	14
B	19	31	48	53	67	90	10	62	40	80

Find the standard deviation of each machine and also comment on the performances of two machines.

Answer: There is less variability in the performance of the **machine B.**

❖ MOMENTS:

- ✓ Moment is a familiar mechanical term which refer to the measure of a force respect to its tendency to provide rotation. OR Moment is the arithmetic mean of the various powers of the deviations of items from their assumed mean or actual mean. If the deviations of the items are taken from the arithmetic mean of the distribution, it is known as central moment.
- ✓ Moments of a Frequency Distribution
  - ➢ The rth moment of a frequency distribution about mean $\overline{X}$ denoted by μ_r is defined by

$$\boldsymbol{\mu_r = \frac{\Sigma f(X-\overline{X})^r}{n}, \text{Where } r = 0, 1, 2, 3, \dots \dots \ (1)}$$

  - ➢ Thus μ_1 stands for first moment about mean, μ_2 stands for second moment about mean and so on are given by

$$\mu_1 = \frac{\Sigma f(X-\overline{X})}{n}, \mu_2 = \frac{\Sigma f(X-\overline{X})^2}{n}, \mu_3 = \frac{\Sigma f(X-\overline{X})^3}{n}, \mu_4 = \frac{\Sigma f(X-\overline{X})^4}{n}$$

- ✓ Moments about any arbitrary origin:
  - ➢ The rth moment about any arbitrary point (value) a denoted by μ'_r is defined by

$$\boldsymbol{\mu'_r = \frac{\Sigma f(x-a)^r}{n} \text{Where } r = 1, 2, 3, \dots \dots .(2)}$$

  - ➢ Particular cases:
    - • Putting $r = 0$ in (2), we get

$$\mu'_0 = \frac{\Sigma f}{n} = 1 \text{ i.e. } \mu'_0 = 1$$

    - • Putting $r = 1$ in (2), we get

$$\mu'_1 = \frac{\Sigma f(x-a)}{n} = \frac{\Sigma fx}{n} - \frac{\Sigma f}{n} \ . \ a \ = \overline{X} - a$$

i.e. $\mu'_1 = \overline{X} - a$ OR $\overline{X} = \mu'_1 + a \dots \dots (3)$

- Putting $r = 2$ in (1)

  $$\mu'_2 = \frac{\Sigma f(X-\overline{X})^2}{n} = \sigma^2 \text{ i.e. } \mu'_2 = \sigma^2 \text{ OR } \sigma = \sqrt{\mu'_2} \;\ldots\ldots (4)$$

❖ RELATION BETWEEN THE MOMENTS ABOUT THE ACTUAL MEAN IN TERMS OF THE MOMENTS ABOUT ANY ARBITRARY VALUE:

- ✓ $\mu_1 = \mu'_1 - \mu'_1 = 0$
- ✓ $\mu_2 = \mu'_2 - (\mu'_1)^2$ ……(5)
- ✓ $\mu_3 = \mu'_3 - 3\mu'_2 \,.\, \mu'_1 + 2(\mu'_1)^3$ ……(6)
- ✓ $\mu_4 = \mu'_4 - 4\mu'_3 \,.\, \mu'_1 + 6\mu'_2 \,.(\mu'_1)^2 - 3(\mu'_1)^4$ ……(7)

❖ MOMENTS ABOUT ZERO

- ✓ The moments about zero are often denoted by $v_{1,}v_{2,}v_{3,}v_4$ are as follows:

  $$v_1 = \frac{\Sigma fx}{n}, v_2 = \frac{\Sigma fx^2}{n}, v_3 = \frac{\Sigma fx^3}{n}, v_4 = \frac{\Sigma fx^4}{n}$$

- ✓ In general , r^{th} moment about zero

  $$\mathbf{v_r = \frac{\Sigma fx^r}{n}} \;\ldots\ldots\; (\mathbf{8})$$

- ✓ Also,
  - ➢ The first moment about zero or $v_1 = a + \mu'_1 = \overline{x}$
  - ➢ The second moment about zero or $v_2 = \mu_2 + (v_1)^2$
  - ➢ The third moment about zero or $v_3 = \mu_3 + 3v_1v_2 - 2{v_1}^3$
  - ➢ The fourth moment about zero or $v_4 = \mu_4 + 4v_1v_3 - 6{v_1}^2v_2 + 3{v_1}^4$

❖ SUMMARY OF HOW MOMENTS HELP IN ANALYZING A FREQUENCY DISTRIBUTION

	Moment	What it measures
	First moment about origin	Mean
	Second moment about the mean	Variance
	Third moment about the mean	Skewness
	Fourth moment about the mean	Kurtosis

❖ SKEWNESS

- ✓ This is a measure of degree of asymmetry of the frequency distribution. The two measures of Skewness namely the Pearson's coefficient of Skewness and quartile coefficients of Skewness are known to us. We introduce another measure of Skewness known as the "coefficient of Skewness based on the third moment "denoted by γ, and is defined as

$$\gamma_1 = \sqrt{\beta_1};\ \text{Where } \beta_1 = \frac{(\mu_3)^2}{(\mu_2)^3}$$

❖ KURTOSIS

- ✓ The measure of peakedness of a distribution (i.e. measure of convexity of a frequency curve) is known as Kurtosis. It is based on fourth moment and is defined as

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

❖ RELATED INFORMATION

- ✓ The greater the value of β_2, the more peaked is the distribution.
- ✓ A frequency distribution for which $\beta_2 = 3$ is a normal curve.
- ✓ When the value of $\beta_2 > 3$, the curve is more peaked than normal curve and the distribution is called leptokurtic.
- ✓ When the value of $\beta_2 < 3$, the curve is less peaked than normal curve and the distribution is called Platykurtic.
- ✓ The normal curve and other curves with $\beta_2 = 3$ are called Mesokurtic.
- ✓ Sometimes $\gamma_2 = \beta_2 - 3$ (is the excess of kurtosis)
- ✓ For a normal distribution $\gamma_2 = 0$.
- ✓ If γ_2 is positive, the curve is leptokurtic and if γ_2 is negative, the curve is Platykurtic.

METHOD – 2: MOMEMNT (SKEWNESS AND KURTOSIS)

C	1	Find mean and first four central moment from the following data: $x_i = 11, 12, 14, 16, 20$ **Answer**: $\mathbf{Mean = 14.60.\ \mu_1 = 0, \mu_2 = 10.24, \mu_3 = 19.1520, \mu_4 = 213.5872}$	

<table>
<tr><td>H</td><td>2</td><td>The quantities of milk (in liters) produced by a dairy farm on ten consecutive days are shown below: 218.2, 199.7, 207.3, 185.4, 213.7, 184.7, 179.5, 194.4, 224.3, 203.5. Evaluate the mean and the first four central moments of the milk yield data (in liters) of the dairy farm.
Answer:
$\mathbf{Mean = 201.07.\ \mu_1 = 0, \mu_2 = 206.3, \mu_3 = 64.2535, \mu_4 = 75447.42}$</td><td></td></tr>
<tr><td>C</td><td>3</td><td>Given the frequency distribution:
<table>
<tr><td>X</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td></tr>
<tr><td>F</td><td>1</td><td>8</td><td>28</td><td>56</td><td>70</td><td>56</td><td>28</td><td>8</td><td>1</td></tr>
</table>
Show that: (a) the distribution is symmetric by establishing the Skewness based on the third moment equal to zero (b) the distribution is platykurtic.
Answer: $\boldsymbol{\gamma_1 = 0\ , \beta_2 = 2.75}$</td><td></td></tr>
<tr><td>H</td><td>4</td><td>Given the frequency distribution: Find the first four central moments and find $\beta_1, \beta_2, \gamma_1, \gamma_2$.
<table>
<tr><td>X</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td></tr>
<tr><td>F</td><td>1</td><td>8</td><td>28</td><td>56</td><td>70</td><td>56</td><td>28</td><td>8</td><td>1</td></tr>
</table>
Answer:
$\mathbf{u_1 = 0, u_2 = 2, u_3 = 0, u_4 = 11}, \boldsymbol{\beta_1 = 0, \beta_2 = 2.75, \gamma_1 = 0, \gamma_2 = -0.25}$</td><td></td></tr>
<tr><td>C</td><td>5</td><td>Calculate moments about (1) assumed mean 25, (2) actual mean, (3) zero.
<table>
<tr><td>Variable</td><td>0 – 10</td><td>10 – 20</td><td>20 – 30</td><td>30 – 40</td></tr>
<tr><td>Frequency</td><td>1</td><td>3</td><td>4</td><td>2</td></tr>
</table>
Answer:
$\boldsymbol{\mu'_1 = -3, \mu'_2 = 90, \mu'_3 = -900, \mu'_4 = 21000}$
$\boldsymbol{\mu_1 = 0, \mu_2 = 81, \mu_3 = -144, \mu_4 = 14817}$
$\mathbf{v_1 = 22, v_2 = 565, v_3 = 798940, v_4 = 69383545}$</td><td></td></tr>
</table>

<table>
<tr>
<td>H</td>
<td>6</td>
<td>Find the first four central moments of the following distribution and comment on the nature of the distribution:
<table>
<tr><td>Class</td><td>100 − 104.9</td><td>105 − 109.9</td><td>110 − 114.9</td><td>115 − 119.9</td><td>120 − 124.9</td></tr>
<tr><td>Frequency</td><td>7</td><td>13</td><td>25</td><td>25</td><td>30</td></tr>
</table>
Answer: $\mu_1 = 0,\ \mu_2 = 38.09,\ \mu_3 = -110.772,\ \mu_4 = 4287.51$</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>7</td>
<td>A continuous r.v. X has a pdf $f(x) = \begin{cases} 3x^2 & ;\ 0 < x \le 1 \\ 0 & ;\ \text{elsewhere} \end{cases}$. Obtain the first four central moments.
Answer: $\mu_1 = 0,\ \mu_2 = \frac{3}{80},\ \mu_3 = -0.006,\ \mu_4 = 0.003$</td>
<td></td>
</tr>
</table>

COEFFICIENT OF CORRELATION:

- Correlation is the relationship that exists between two or more variables. Two variables are said to be correlated if a change in one variable affects a change in the other variable. Such a data connecting two variables is called bivariate data.
- When two variables are correlated with each other, it is important to know the amount or extent of correlation between them. The numerical measure of correlation of degree of relationship existing between two variables is called the coefficient of correlation and is denoted by r and it is always lies between -1 and 1.
  - When $r = 1$, it represents Perfect Direct or Positive Correlation.
  - When $r = -1$ it represents Perfect Inverse or Negative Correlation.
  - When $r = 0$, there is No Linear Correlation or it shows Absence Of Correlation.
  - When the value of r is ± 0.9 or ± 0.8 etc. it shows high degree of relationship between the variables and when r is small say ± 0.2 or ± 0.1 etc, it shows low degree of correlation.

TYPES OF CORRELATIONS:

- Positive and negative correlations
- Simple and multiple correlations
- Partial and total correlations
- Linear and nonlinear correlations

(1). Positive and Negative Correlations:

- Depending on the variation in the variables, correlation may be positive or negative.
- Positive Correlation:

  If both the variables vary in the same direction, the correlation is said to be positive. In the other words, if the value of one variable increases, the value of the other variable also increases, or, if the value of one variable decreases, the value of the other variable also decreases, e.g., the correlation between heights and weights of group of persons is a positive correlation.
- Negative Correlation:

  If both the variables vary in the opposite direction, correlation is said to be negative. In other words, if the value of one variable increases, the value of the other variable also decreases, or, if the value of one variable decreases, the value of the other variable also increases, e.g., the correlation between the price and demand of a commodity is a negative correlation.

(2). Simple and Multiple Correlation:

- Depending upon the study of the number of variables, correlation may be simple or multiple.
- Simple Correlation

  When only two variables are studied, the relationship is described as simple correlation, e.g., the quantity of money and price level, demand and price, etc.
- Multiple correlation

  When more than two variables are studied, the relationship is described as multiple correlation, e.g., relationship of price, demand, and supply of a commodity.

(3). Partial and Total Correlation:

- Multiple correlation may be either partial or total.
- Partial Correlation:

  When more than two variables are studied excluding some other variables, the relationship is termed as partial correlation.
- Total Correlation:

  When more than two variables are studied, without excluding any variables, the relationship is termed as total correlation.

(4). Linear and Nonlinear Correlation:

- Depending upon the ratio of change between two variables, the correlation may be linear or nonlinear.
- Linear Correlation:

  If the ratio of change between two variables is constant, the correlation is said to be linear.
- Nonlinear Correlation:

  If the ratio of change between two variables is not constant, the correlation is said to be nonlinear.

❖ METHODS OF STUDYING CORRELATION:

✓ There are two different methods of studying correlation,

(1). Graphic methods

- scatter diagram
- Simple graph

(2). Mathematical methods

- Karl Pearson's coefficient of correlation
- Spearman's rank coefficient of correlation

❖ KARL PEARSON'S PRODUCT MOMENT METHOD:

✓ This is most popular and widely used mathematical method. In this method the degree of correlation between two variables can be measured by the coefficient of correlation and it gives not only magnitude (degree) of correlation but also its direction.

✓ Let $(x_{1,}y_1), (x_{2,}y_2), \ldots \ldots \ldots (x_{n,}y_n)$ be n pairs of observations of two variables X and Y, then coefficients of correlation (r) between X and Y is defined by

$$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{n\,\sigma_x\,.\,\sigma_y}$$

OR $$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2}\,\sqrt{\Sigma(Y-\overline{Y})^2}}$$

OR $$r = \frac{n\,\Sigma\,x\,y-(\Sigma\,x)(\Sigma\,y)}{\sqrt{n\,\Sigma\,x^2-(\Sigma\,x)^2}\sqrt{n\Sigma\,y^2-(\Sigma\,y)^2}} \text{(small no. of itemes in two variables)}$$

$$\textbf{OR}\ \ r = \frac{n\,\Sigma\,dx\,dy - (\Sigma\,dx)\,(\Sigma\,dy)}{\sqrt{n\,\Sigma\,dx^2 - (\Sigma\,dx)^2}\sqrt{n\,\Sigma dy^2 - (\Sigma\,dy)^2}} \text{(deviaton by assumed mean)}$$

$$\textbf{OR}\ \ r = \frac{\text{Cov}(X, Y)}{\sigma_x \,.\, \sigma_y}$$

- ✓ Where,
  - ➢ $\text{Cov}\,(X, Y) = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{n}$
  - ➢ $n =$ Number of X, Y pairs
  - ➢ $\bar{X} =$ Mean of $X = \frac{\sum X}{n}$
  - ➢ $\bar{Y} =$ Mean of $Y = \frac{\sum Y}{n}$
  - ➢ $\sigma_x =$ standard deviation of $X = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n}}$
  - ➢ $\sigma_y =$ standard deviation of $Y = \sqrt{\frac{\Sigma(Y-\bar{Y})^2}{n}}$
  - ➢ $dx = X - A$ & $dy = Y - B$. A & B being assumed means of X and Y respectively.

❖ REGRESSION EQUATIONS:

- ✓ The algebraic expressions of the regression lines are called Regression equations. Since there are two regression lines, there are two regression equations.
- ✓ Using method of least squares we have obtained the regression equation of y on x as $y = a + bx$ and that of x on y as $x = a + by$. The values of a and b depends on the means, the standard deviations and coefficient of correlation between the two variables.
- ✓ These equations are written as follows:

$$\mathbf{y - \bar{y} = b_{yx}(x - \bar{x})} \qquad \textbf{(y on x)}$$
$$\mathbf{x - \bar{x} = b_{xy}(y - \bar{y})} \qquad \textbf{(x on y)}$$

- ✓ Where,
  - ➢ $\bar{x} =$ Mean of $x = \frac{\sum x}{n}$
  - ➢ $\bar{y} =$Mean of $y = \frac{\sum y}{n}$
  - ➢ $b_{yx} = r.\frac{\sigma_y}{\sigma_x}$ (regression coefficient)
  - ➢ $b_{xy} = r.\frac{\sigma_x}{\sigma_y}$ (regression coefficient)
  - ➢ $\sigma_x =$ Standard deviation of x

- σ_y = Standard deviation of y
- $\mathbf{r}$ = Correlation coefficient between x & y

❖ FORMULA FOR COMPUTATIONS OF REGRESSIONS COEFFICIENTS:

- ✓ The following formulae are used for calculation of $\mathbf{b_{yx}}$ and $\mathbf{b_{xy}}$,
  - When σ_x, σ_y and r are given
    - $b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}$ & $b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y}$
  - When $n, \sum x, \sum y, \sum x^2, \sum y^2$ and $\sum xy$ are given
    - $b_{yx} = \frac{n\sum xy-(\sum x)(\sum y)}{n\sum x^2-(\sum x)^2}$ & $b_{xy} = \frac{n\sum xy-(\sum x)(\sum y)}{n\sum y^2-(\sum y)^2}$
  - When the deviations are taken from their means
    - $b_{yx} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sum(X-\bar{X})^2}$ & $b_{xy} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sum(Y-\bar{Y})^2}$
  - When the deviations are taken from assumed numbers
    - $b_{yx} = \frac{n\sum dxdy-(\sum dx)(\sum dy)}{n\sum dx^2-(\sum dx)^2}$ & $b_{xy} = \frac{n\sum dxdy-(\sum dx)(\sum dy)}{n\sum dy^2-(\sum dy)^2}$

      ; Where $\mathbf{dx = x - A}, \mathbf{dy = y - B}$. **A** and **B** are any assumed numbers.
  - When the values of covariance and variances are given
    - $b_{yx} = \frac{Cov(X,Y)}{\sigma_x^2}$ & $b_{xy} = \frac{Cov(X,Y)}{\sigma_y^2}$; where $x = (X - \bar{X}), y = (Y - \bar{Y})$

❖ PROPERTIES OF REGRESSION COEFFICIENTS:

- ✓ The geometric mean ($\mathbf{r}$) between two regression coefficients is given by

  i.e. $\mathbf{r = \sqrt{b_{yx} \times b_{xy}}}$
- ✓ Both the regression coefficients will have the same sign.

  i.e. they are either both positive or both negative.
- ✓ The values of both $\mathbf{b_{xy}}$ and $\mathbf{b_{yx}}$ individually cannot be more than 1.
- ✓ The Sign of r is same as of regression coefficients, if $r < 0$ then $b_{yx} < 0$ & $b_{xy} < 0$.

METHOD – 3: CORRELATION AND REGRESSION LINE

<table>
<tr><td>H</td><td>1</td><td>Calculate the co-efficient of correlation between the given series of data for x and y in the following table:
<table>
<tr><td>X</td><td>54</td><td>57</td><td>55</td><td>57</td><td>56</td><td>52</td><td>59</td></tr>
<tr><td>Y</td><td>36</td><td>35</td><td>32</td><td>34</td><td>36</td><td>38</td><td>35</td></tr>
</table>
Answer: $\mathbf{r = -0.4575}$</td><td></td></tr>
<tr><td>H</td><td>2</td><td>Calculate correlation coefficient from following data.
<table>
<tr><td>X</td><td>1100</td><td>1200</td><td>1300</td><td>1400</td><td>1500</td><td>1600</td><td>1700</td><td>1800</td><td>1900</td><td>2000</td></tr>
<tr><td>Y</td><td>0.30</td><td>0.29</td><td>0.29</td><td>0.25</td><td>0.24</td><td>0.24</td><td>0.24</td><td>0.29</td><td>0.18</td><td>0.15</td></tr>
</table>
Answer: $\mathbf{r = -0.7906}$</td><td></td></tr>
<tr><td>C</td><td>3</td><td>Determine the coefficient of correlation $n = 10, \bar{x} = 5.5; \bar{y} = 4; \sum x^2 = 385;$ $\sum y^2 = 192\,; \sum (x+y)^2 = 947.$
Answer: $\mathbf{r = -0.681}$</td><td></td></tr>
<tr><td>H</td><td>4</td><td>Determine the coefficient of correlation $n = 8, \bar{x} = 0.5; \bar{y} = 0.5; \sum x^2 = 44;$ $\sum y^2 = 44\,; \sum xy = -40.$
Answer: $\mathbf{r = -1}$</td><td></td></tr>
<tr><td>C</td><td>5</td><td>Find r_{xy} from given data: $n = 10, \sum (x-\bar{x})(y-\bar{y}) = 1650,\ \ \sigma_x^2 = 196,\ \ \sigma_y^2 = 225.$
Answer: $\mathbf{r_{xy} = 0.7857}$</td><td></td></tr>
<tr><td>H</td><td>6</td><td>Find $r_{\,xy}$ from given data: $n = 10, \sum (x-\bar{x})(y-\bar{y}) = 66,\ \ \sigma_x = 5.4,\ \ \sigma_y = 6.2.$
Answer: $\mathbf{r_{xy} = 0.197}$</td><td></td></tr>
<tr><td>H</td><td>7</td><td>Explain Co- relation, co-relation types, co-relation co-efficient. Also state the method to find correlation between two variables. Find the correlation coefficient between the serum diastolic B.P. and serum cholesterol levels of 10 randomly selected data of 10 persons.
<table>
<tr><td>Person</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td></tr>
<tr><td>Choles.</td><td>307</td><td>259</td><td>341</td><td>317</td><td>274</td><td>416</td><td>267</td><td>320</td><td>274</td><td>336</td></tr>
<tr><td>B.P.</td><td>80</td><td>75</td><td>90</td><td>74</td><td>75</td><td>110</td><td>70</td><td>85</td><td>88</td><td>78</td></tr>
</table>
Answer: $\mathbf{r = 0.8087}$</td><td></td></tr>
</table>

T **8** Calculate Karl-Pearson’s correlation coefficient between age and playing habits from the data given below.

Age	20	21	22	23	24	25
No. of students	500	400	300	240	200	160
Regular players	400	300	180	96	60	24

Answer: $\mathbf{r = -0.9912}$

C **9** From the following data calculate two equation of line of regression:

	X	Y
Mean	60	67.5
Standard deviation	15	13.5

Correlation coefficient between X & Y is 0.50. Also estimate the value of Y for $X = 72$ using the appropriate regression equation.

Answer: $\mathbf{Y = 40.5 + 0.45X, X = 22.47 + 0.556Y, Y(72) = 72.9}$

H **10** The following values are available for the variable x & y.

$n = 10, \sum x = 30, \sum y = 40, \sum x^2 = 222, \sum y^2 = 985, \sum xy = 384.$

Obtain: Two regression equation and Correlation coefficient.

Answer: $\mathbf{Y = 2X - 2, X = 0.32Y + 1.72, r = 0.8}$

C **11** Find the lines of regression of Y on X if $n = 9, \sum x = 30.3, \sum y = 91.1, \sum xy = 345.09$, & $\sum x^2 = 115.11$. Also find value of Y(1.5) & Y(5.0).

Answer: $\mathbf{Y = 2.93X + 0.2568, Y(1.5) = 4.6523, Y(5.0) = 14.9083}$

C **12** The amount of chemical compound , which were dissolved in 100 grams of water at various temperatures , x were recorded as:

x(°C)	15	15	30	30	45	45	60	60
y(grams)	12	10	25	21	31	33	44	39

Find the line of regression of y on x & estimate the amount of chemical that will dissolve in 100 grams of water at 50°C.

Answer: $\mathbf{Y = 0.67X + 1.75, Y(50^\circ) = 35.25}$

C **13** Obtain the two lines of regression for the following data:

X	190	240	250	300	310	335	300
Y	5	10	15	20	20	30	30

Answer: $\mathbf{X = 184.867 + 4.8533Y, Y = -28.6233 + 0.1716X}$

H **14** Calculate the coefficient of correlation & obtain lines of regression for:

X	1	2	3	4	5	6	7	8	9
Y	9	8	10	12	11	13	14	16	15

Answer: $\mathbf{r = 0.95\ , X = -6.4 + 0.95Y\ , Y = 7.25 + 0.95X}$

H **15** From the following obtain two regression lines & the correlation coefficients.

X	100	98	78	85	110	93	80
Y	85	90	70	72	95	81	74

Answer: $\mathbf{r = 0.96\ , X = -2.77 + 1.17Y, Y = 8.78 + 0.79X}$

H **16** Following data give experience of machine operators & their performances rating as given by the number of good parts turned out per 100 piece.

Operator	1	2	3	4	5	6
Performance rating(X)	23	43	53	63	73	83
Experience(Y)	5	6	7	8	9	10

(a) Calculate the regression line of performing rating on experience and also estimate the probable performance if an operator has 11 years' experience.

(b) Define regression coefficients and give its properties.

Answer: $\mathbf{X = 11.428\ Y - 29.38,\ 96.33}$

T **17** A study of the amount of daily rainfall (x) and the quantity of air pollution removed produced (y) the following data:

x (0.01 cm)	4.3	4.5	5.9	5.6	6.1	5.2	3.8	2.1	7.5
y (μg/m^3)	126	121	116	118	114	118	132	141	108

(a) Find the equation of the regression line to predict the particulate removed from the amount of Daily rainfall.

(b) Find amount of particulate removed when daily rainfall is $x = 4.8$ units.

Answer: $\mathbf{y = 153.175 - 6.324x\ ; y(4.8) = 122.8198}$

❖ SPEARMAN'S RANK CORRELATION METHOD:

- ✓ Rank correlation is based on the rank or the order of the variables and not on the magnitude of the variables. Here, the individuals are arranged in order of proficiency.
- ✓ If the ranks are assigned to the individuals range from 1 to n, then the correlation coefficient between two series of ranks is called rank correlation coefficient.
- ✓ Edward Spearman's formula for rank coefficient of correlation(R) is given by

$$\mathbf{R = 1 - \frac{6\sum d^2}{n(n^2-1)}}$$

- ✓ Where,
  - ➢ $d =$ difference between the ranks R_1 and R_2 given by two judges
  - ➢ $n =$ number of pairs
- ✓ For example, we cannot measure beauty and intelligence quantitatively. It is possible to rank the individuals in order.
- ✓ The above formula is used when ranks are not repeated.

❖ TIED RANKS:

- ✓ If there is a tie between two or more individuals' ranks, the rank is divided among equal individuals.
- ✓ **EXAMPLES:**
  - ➢ If two items have fourth rank, the 4th and 5th rank is divided between them equally and is given as $\frac{4+5}{2} = 4.5^{\text{th}}$ rank to each of them.
  - ➢ If three items have the same 4th rank, each of them is given $\frac{4+5+6}{3} = 5^{\text{th}}$ rank.
- ✓ As a result of this, the following adjustment or correlation is made in the rank correlation formula.
- ✓ If m is the number of item having equal ranks then the factor $\frac{1}{12}(m^3 - m)$ is added to $\sum d^2$. If there are more than one cases of this types, this factor is added corresponding to each case.

$$\mathbf{r = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}({m_1}^3 - m_1) + \frac{1}{12}({m_2}^3 - m_2) + \cdots\right]}{n(n^2-1)}}$$

METHOD – 4: RANK CORRELATION

C **1** Two Judges in a beauty contest rank the 12 contestants as follows:

X	1	2	3	4	5	6	7	8	9	10	11	12
Y	12	9	6	10	3	5	4	7	8	2	11	1

What degree of agreement is there between the Judges?

Answer: $\boldsymbol{\rho = -0.4545}$

H **2** The participant in contest are raked by two judges as follows:

X	1	3	7	5	4	6	2	10	9	8
Y	3	1	4	5	6	9	7	8	10	2

Calculate the rank correlation coefficient.

Answer: 0.418

C **3** Calculate coefficient of correlation by spearman's method from the following.

Sales	45	56	39	54	45	40	56	60	30	36
Cost	40	36	30	44	36	32	45	42	20	36

Answer: $\boldsymbol{\rho = 0.7636}$

H **4** Obtain the rank correlation coefficient for the following data:

X	68	64	75	50	64	80	75	40	55	64
y	62	58	68	45	81	60	68	48	50	70

Answer: $\boldsymbol{\rho = 0.545}$

C **5** In a college, IT department has arranged one competition for IT students to develop an efficient program to solve a problem. Ten students took part in the competition and ranked by two judges given in the following table. Find the degree of agreement between the two judges using Rank correlation coefficient.

1st judge	3	5	8	4	7	10	2	1	6	9
2nd judge	6	4	9	8	1	2	3	10	5	7

Answer: $\boldsymbol{\rho = -0.2970}$

H **6** 9 students secured the following percentage of marks in math and chemistry. Find the rank correlation coefficient and comment on its value.

Roll no.	1	2	3	4	5	6	7	8	9
Marks in Math.	78	36	98	25	75	82	90	62	65
Marks in Chem.	84	51	91	60	68	62	86	58	53

Answer: $\boldsymbol{\rho = 0.8333}$

T **7** The competitions in a beauty contest are ranked by three judges:

1st judge	1	5	4	8	9	6	10	7	3	2
2nd judge	4	8	7	6	5	9	10	3	2	1
3rd judge	6	7	8	1	5	10	9	2	3	4

Use rank correlation coefficient method to discuss which pair of judges has nearest approach to beauty.

Answer: $\boldsymbol{\rho_1 = 0.5515\,,\rho_2 = 0.7333\,,\rho_3 = 0.05454}$

UNIT-4 » APPLIED STATISTICS

❖ INTRODUCTION:

- ✓ Many problems in engineering required that we decide which of two competing claims for statements about parameter is true. Statements are called Hypotheses, and the decision making procedure is called hypotheses testing. This is one of the most useful aspects of statistical inference, because many type of decision making problems, tests or experiments in the engineering world can be formulated as hypotheses testing problems.

❖ POPULATION OR UNIVERSE

- ✓ An aggregate of objects (animate or inanimate) under study is called population or universe. It is thus a collection of individuals or of their attributes (qualities) or of results of operations which can be numerically specified.
- ✓ A universe containing a finite number of individuals or members is called a finite universe. For example, the universe of the weights of students in a particular class or the universe of smokes in Rothay district.
- ✓ A universe with infinite number of members is known as an infinite universe. For example, the universe of pressures at various points in the atmosphere.
- ✓ In some cases, we may be even ignorant whether or not a particular universe is infinite, e.g., the universe of stars.
- ✓ The universe of concrete objects is an existent universe. The collection of all possible ways in which a specified event can happen is called a hypothetical universe. The universe of heads and tails obtained by tossing an infinite number of times is a hypothetical one.

❖ SAMPLING

- ✓ A finite sub-set of a universe or population is called a sample. A sample is thus a small portion of the universe. The number of individuals in a sample is called the sample size. The process of selecting a sample from a universe is called sampling.
- ✓ The theory of sampling is a study of relationship between a population and samples drawn from the population. The fundamental object of sampling is to get as much information as possible of the whole universe by examining only a part of it.
- ✓ Sampling is quite often used in our day-to-day practical life. For example, in a shop we assess the quality of sugar, rice or any commodity by taking only a handful of it from the bag and then decide whether to purchase it or not.

❖ TEST OF SIGNIFICANCE

- ✓ An important aspect of the sampling theory is to study the test of significance. Which will enable us to decide, on the basis of the results of the sample. Whether
  - ➢ The deviation between observed sample statistic and the hypothetical parameter value
  - ➢ The deviation between two samples statistics is significant of might be attributed due to chance or the fluctuations of the sampling.
- ✓ For applying the tests of significance, we first set up a hypothesis which is a definite statement about the population parameter called null hypothesis denoted by H_0.
- ✓ Any hypothesis which is complementary to the null hypothesis (H_0) is called an alternative hypothesis denoted by H_1.
- ✓ For example, if we want to test the null hypothesis that the population has a specified mean μ_0, then we have $H_0 : \mu = \mu_0$
- ✓ Alternative hypothesis will be
  - ➢ $H_1 : \mu \neq \mu_0$ ($\mu > \mu_0$ or $\mu < \mu_0$) (Two tailed alternative hypothesis).
  - ➢ $H_1 : \mu > \mu_0$ (Right tailed alternative hypothesis or single tailed).
  - ➢ $H_1 : \mu > \mu_0$ (Left tailed alternative hypothesis or single tailed).
- ✓ Hence alternative hypothesis helps to know whether the test is two tailed or one tailed test.

❖ STANDARD ERROR

- ✓ The standard deviation of the sampling distribution of a statistic is known as the standard error (S.E.).
- ✓ It plays an important role in the theory of large samples and it forms a basis of testing of hypothesis. If t is any statistic, for large sample.
- ✓ $z = \frac{t-E(t)}{S.E(t)}$ Is normally distributed with mean 0and variance unity.
- ✓ For large sample, the standard errors of some of the well-known statistic are listed below:

No.	Statistic	Standard error
1	$\bar{x}$	$\frac{\sigma}{\sqrt{n}}$
2	S	$\sqrt{\frac{\sigma^2}{2n}}$

3	Difference of two sample means $\overline{x_1} - \overline{x_2}$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
4	Difference of two sample standard deviation $s_1 - s_2$	$\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$
5	Difference of two sample proportions $p_1 - P_2$	$\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$
6	Observed sample proportion p	$\sqrt{\frac{PQ}{n}}$

❖ ERRORS IN SAMPLING

- ✓ The main aim of the sampling theory is to draw a valid conclusion about the population parameters. On the basis of the same results. In doing this we may commit the following two type of errors:
  - **TYPE I ERROR:** When H_0 is true, we may reject it.

    $$P(\text{Reject } H_0 \text{ when it is true}) = P\left(\text{Reject}\frac{H_0}{H_0}\right) = \alpha$$

    Where, α is called the size of the type I error also referred to as product's risk.
  - **TYPE II ERROR:** When H_0 is wrong we may accept it.

    $$P(\text{Accept } H_0 \text{ when it is wrong}) = P\left(\text{Accept}\frac{H_0}{H_1}\right) = \beta$$

    Where β is called the size of the type II error, also referred to as consumer's risk.

❖ STEPS FOR TESTING OF STATISTICAL HYPOTHESIS:

- ✓ **Step 1:** Null hypothesis.
  - Set up H_0 is clear terms.
- ✓ **Step 2:** Alternative hypothesis.
  - Set up H_1 , so that we could decide whether we should use one-tailed or two-tailed test.
- ✓ **Step 3:** Level of significance.
  - Select appropriate level of significance in advance depending on reality of estimates.
- ✓ **Step 4:** Test statistic.
  - Compute the test statistic $z = \frac{t-E(t)}{S.E(t)}$ under the null hypothesis.
- ✓ **Step 5:** Conclusion.

- Compare the computed value of z with critical value z_α at the level of significance (α).
  - If $|z| > z_\alpha$, we reject H_0 and conclude that there is significant difference.
  - If $|z| < z_\alpha$, we accept H_0 and conclude that there is no significant difference.

❖ TEST OF SIGNIFICANCE FOR LARGE SAMPLES

(1). TESTING OF SIGNIFICANCE FOR SINGLE PROPORTION:

- This test is used to find the significant difference between proportion of the sample and the population.
- Let X be the number of successes in n independent trials with constant probability P of success for each trial.

$$E(X) = nP\ ; V(X) = nPQ\ \ ; Q = 1 - P = \text{Probability of failure.}$$

$$E(p) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{np}{n} = p\ ; E(p) = p$$

$$V(p) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}V(x) = \frac{1(PQ)}{n} = \frac{PQ}{n}$$

$$S.E(p) = \sqrt{\frac{PQ}{n}}\ ; z = \frac{p - E(p)}{S.E(p)} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \sim N(0,1)$$

- This z is called test statistic which is used to test the significant difference of sample and population proportion.
- The probable limit for the observed proportion of successes are $p \pm z_\alpha\sqrt{\frac{PQ}{n}}$, where z_α is the significant value at level of significance α.
- If p is not known, the limits for proportion in the population are $p \pm z_\alpha\sqrt{\frac{pq}{n}}$, $q = 1 - p$.
- If α is not known, we can take safely 3σ limits.
- Hence, confidence limits for observed proportion p are $p \pm 3\sqrt{\frac{PQ}{n}}$.
- The confidence limits for the population proportion p are $p \pm \sqrt{\frac{pq}{n}}$.

METHOD – 1: TESTING OF SIGNIFICANCE FOR SINGLE PROPORTION

C	1	A political party claims that 45% of the voters in an election district prefer its candidate. A sample of 200 voters include 80 who prefer this candidate. Test if the claim is valid at the 5% significance level. **Answer: The party's claim might be valid.**	
H	2	In a sample of 400 parts manufactured by a factory, the number of defective parts found to be 30. The company, however, claims that only 5% of their product is defective. Is the claim tenable? (Take level of significance 5%) **Answer: The claim of manufacturer is not tenable.**	
C	3	A certain cubical die was thrown 9000 times and 5 or 6 was obtained 3240 times. On the assumption of certain throwing, do the data indicate an unbiased die? **Answer: The die is unbiased.**	
H	4	A bag contains defective article, the exact number of which is not known. A sample of 100 from the bag gives 10 defective articles. Find the limits for the proportion of defectivea articles in the bag. **Answer: 95% confidence limits for defective article in bag are $(\mathbf{0.1588}, \mathbf{0.0412})$**	

(2). TESTING OF DIFFERENCE BETWEEN PROPORTION

- ✓ Consider two samples X_1 and X_2 of sizes n_1 and n_2 respectively taken from two different population. To test the significance of the difference between sample proportion p_1 & p_2.
- ✓ The test statistic under the null hypothesis H_0, that there is no significant difference between the two sample proportions, we have

$$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \text{where } P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \text{ and } Q = 1 - P$$

METHOD – 2: TESTING OF DIFFERENCE BETWEEN PROPORTION

C	**1**	In a certain city A,100 men in a sample of 400 are found to be smokers. In another city B, 300 men in a sample of 800 are found to be smokers. Does this indicate that there is greater proportion of smokers in B than in A?(use level of significant 5%) **Answer: The proportion of smokers is greater in the second city B than in A.**	
H	**2**	Before an increase in excise duty on tea,800 people out of a sample of 1000 persons were found to be tea drinkers. After an increase in the duty, 800 persons were known to be tea drinkers in a sample of 1200 persons. Do you think that there has been a significant decrease in the consumption of tea after the increase in the excise duty? **Answer: There is a significant decrease in the consumption of tea due to increase in excise duty.**	
C	**3**	A question in a true-false is considered to be smart if it discriminates between intelligent person (IP) and average person (AP).Suppose 205 out of 250 IP's and 137 out of 250 AP's answer a quiz question correctly. Test of 0.01 level pf significance whether for the given question, proportion of correct answers can be expected to be at least 15% higher among IP's than among the AP's. **Answer: Proportion of correct answer by IP's is 15% more than those by AP's.**	
H	**4**	500 Articles from a factory are examined and found to be 2% defective.800 Similar articles from a second factory are found to have only 1.5% defective. Can it reasonably be concluded that the product of the first factory are inferior those of second? **Answer: Products do not defer in quality.**	

(3). TESTING OF SIGNIFICANCE FOR SINGLE MEAN

- ✓ To test whether the difference between sample mean and population mean is significant or not.
- ✓ Let $X_1, X_2, \ldots \ldots X_n$ be a random sample of size n from a large population $X_1, X_2, \ldots \ldots X_N$ of size N with mean μ and variance σ^2. Therefore the standard error of mean of a random sample of size n from a population with variance σ^2 is $\frac{\sigma}{\sqrt{n}}$.

- ✓ To test whether given sample of size n has been drawn from a population with mean μ i.e., to test whether the difference between the sample mean and population mean is significant or not. Under the null hypothesis that there is no difference between the sample mean and population mean.
- ✓ The test statistic is $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$; where σ is the standard deviation of the population.
- ✓ If σ is not known, we use test statistic $z = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$, where s is standard deviation of the sample.
- ✓ If the level of significance is α and z_α is the critical value $-z_\alpha < |z| = \left|\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| < z_\alpha$
- ✓ The limit of the population mean μ are given by $\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$.
- ✓ **Confidence limits:**
  - ➢ At 5% of level of significance, 95% confidence limits are $\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$.
  - ➢ At 1% of level of significance, 99% confidence limits are $\bar{x} - 2.58\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.58\frac{\sigma}{\sqrt{n}}$.

METHOD – 3: TESTING OF SIGNIFICANCE FOR SINGLE MEAN

C	1	Let X be the length of a life of certain computer is approximately normally distributed with mean 800 days and standard deviation 40 days. If a random sample of 30 computers has an average life 788 days, test the null hypothesis that $\mu \neq 800$ days at (a) 0.5 % (b) 15% level of significance. **Answer: (a) Accept null hypothesis, (b) Reject null hypothesis.**	
H	2	The mean weight obtained from a random sample of size 100 is 64 gms. The S.D. of the wight distribution of the population is 3 gms.Test the statement that the mean weight of the population is 67 gms. At 5% level of significance. Also set up 99% confidence limits of the mean weight of the population. **Answer: The sample is not drawn from population with mean 67,** $(\mathbf{63.226, 64.774})$	
C	3	A college claims that its average class size is 35 students. A random sample of 64 students class has a mean of 37 with a standard deviation of 6. Test at the $\alpha = 0.05$ level of significance if the claimed value is too low. **Answer: The true mean class size is likely more than 35.**	

H	4	Sugar is packed in bags by an automation machine with mean contents of bags as 1.000 kg. A random sample of 36 bags is selected and mean mass has been found to be 1.003 kg.If a S.D. of 0.01 kg is acceptable on all the bags being packed, determine on the basis of sample test whether the machine requires adjustment.(take level of significance 5%). **Answer: The machine does not require any adjustment.**	

(4). TESTING OF SIGNIFICANCE FOR DIFFERENCE OF MEANS OF TWO LARGE SAMPLES

✓ Let $\overline{x_1}$ be the mean of a sample of size n_1 from a population with mean μ_1 and variance σ_1^2. Let $\overline{x_2}$ be the mean of an independent sample of size n_2 from another population with mean μ_2 and variance σ_2^2.The test statistic is given by

$$z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

✓ Under the null hypothesis that the samples are drawn from the same population where $\sigma_1 = \sigma_2 = \sigma$ i.e., $\mu_1 = \mu_2$ the test statistic is given by

$$z = \frac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

✓ If σ_1, σ_2 are not known and $\sigma_1 \neq \sigma_2$ the test statistic in this case is

$$z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

✓ If σ is not known and $\sigma_1 = \sigma_2$ we use $\sigma^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$ to calculate σ ;

$$z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

METHOD: 4 TESTING OF SIGNIFICANCE FOR DIFFERENCE OF MEANS

C	1	In a random sample of 100 light bulbs manufactured by a company A, the mean lifetime of light bulb is 1190 hours with standard deviation of 90 hours. Also, in a random sample of 75 light bulbs manufactured by company B, the mean lifetime of light bulb is 1230 hours with standard deviation of 120 hours. Is there a difference between the mean lifetimes of the two brands of light bulbs at a significance level of **(a)** 0.05 **(b)** 0.01? **Answer:** **(a)** there is a difference between the mean lifetimes **(b)** there is no difference between the mean lifetimes	
H	2	A company A manufactured tube lights and claims that its tube lights are superior than its main competitor company B. The study showed that a sample of 40 tube lights manufactured by company A has a mean lifetime of 647 hours of continuous use with a standard deviation of 27 hours, while a samle of 40 tube lights manufactured by company had a mean lifetime 638 hours of continuous use with a standard deviation of 31 hours. Does this substantiate the claim of company A that their tube lights are superior than manufactured by company B at **(a)** 0.05 **(b)** 0.01 level of significance? **Answer:** **(a)** There is no difference between tube lights manufactured by two companies A and B. **(b)** There is no difference between tube lights manufactured by two companies A and B.	
C	3	For sample I, $n_1 = 1000, \sum x = 49{,}000, \sum(x - \bar{x})^2 = 7{,}84{,}000$. For sample II, $n_2 = 1500, \sum x = 70{,}500, \sum(x - \bar{x})^2 = 24{,}00{,}000$. Discuss the significance of the difference of the sample means. **Answer: There is no significance difference between the sample means.**	

H **4** A company claims that alloying reduces resistance of electric wire by more than 0.050 ohm. To test this claim samples of 32 standard wire and alloyed wire are tested yielding the following results.

Type of wire	Mean resistance (ohms)	S.D.(s) (ohms)
Standard	0.136	0.004
Alloyed	0.083	0.005

At the 0.05 level of significance, does this support the claim?

Answer: The data substantiate support the claim.

(5). TEST OF SIGNIFICANCE FOR THE DIFFERENCE OF STANDARD DEVIATION

- ✓ If s_1 and s_2 are the standard deviations of two independent samples then under the null hypothesis $H_0: \sigma_1 = \sigma_2$, i.e., the sample standard deviation don't differ significantly, the static

$$z = \frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}} \quad (Where, \sigma_1 \ \ and \ \ \sigma_2 \ \ are\ population\ standard\ deviations)$$

- ✓ When population standard deviations are not known then

$$z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}},$$

METHOD: 5 TEST OF SIGNIFICANCE FOR THE DIFFERENCE OF STANDARD DEVIATION

C **1** Random samples drawn from two countries gave the following data relating to the heights of adult males.

	Country A	Country b
Mean height (inches)	67.42	67.25
Standard deviation	2.58	2.5
Number in samples	1000	1200

Is the difference between the standard deviation significant?

Answer: The sample standard deviations do not differ significantly.

<table>
<tr><td>H</td><td>2</td><td>Intelligence test of two groups of boys and girls gives the following results:
<table>
<tr><th></th><th>Mean</th><th>S.D.</th><th>N</th></tr>
<tr><td>Girls</td><td>84</td><td>10</td><td>121</td></tr>
<tr><td>Boys</td><td>81</td><td>12</td><td>81</td></tr>
</table>
Is the difference between the standard deviations significant?
Answer: The sample standard deviations do not differ significantly.</td><td></td></tr>
<tr><td>C</td><td>3</td><td>The mean yield of two plots and their variability are as given below:
<table>
<tr><th></th><th>40 plots</th><th>60 plots</th></tr>
<tr><td>Mean</td><td>1258</td><td>1243</td></tr>
<tr><td>S.D.</td><td>34</td><td>28</td></tr>
</table>
Check whether the difference in the variability in yields is significant.
Answer: The sample standard deviations do not differ significantly.</td><td></td></tr>
<tr><td>H</td><td>4</td><td>The yield of wheat in a random sample of 1000 farms in a certain area has a S.D. of 192 kg. Another random sample of 1000 farms gives a S.D. of 224 kg. Are the S. Ds significantly different?
Answer: The S. Ds are significantly different</td><td></td></tr>
</table>

❖ TEST OF SIGNIFICANCE FOR SMALL SAMPLES:

(1). t-TEST OF SIGNIFICACE OF THE MEAN:

- ✓ To test whether the mean of a sample drawn from a normal population deviates significantly from a stated value when variance of the population is unknown.
- ✓ H_0 : There is no significant difference between the sample mean $\overline{X}$ and the population mean μ i. e., we use the static

$$t = \frac{\overline{X}-\mu}{s/\sqrt{n}} \qquad \{ Where\ \bar{X}\ is\ mean\ of\ the\ sample\ and\ s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2 \}$$

- ✓ This test static is known as one sample t-test.

METHOD: 6 T-TEST OF SIGNIFICACE OF THE MEAN

C	1	A random sample of size 16 has 53 as mean.The sum of squares of the derivation from mean is 135.Can this sample be regarded as taken from the population having 56 as mean? **Answer: The sample mean has not come from a population mean.**	
H	2	A machine is designed to produce insulting washers for electrical devices of average thickness of 0.025 cm. A random sample of 10 washers was found to have an average thickness of 0.024 cm wih S.D. of 0.002 cm. Test the significance of the deviation.(value of t for 9 degree of freedom at 5% level of significance is 2.262) **Answer: There is no significance difference between population mean and sample mean.**	
C	3	Ten individuals were chosen random from a normal population and their heights were found to be in inches 63,63,66,67,68,69,70,70,71 and 71.Test the hypothesis that the mean height of the population is 66 inches. (Take level of significance 0.05) **Answer: There is no significance difference between population mean and sample mean.**	
H	4	The 9 items of a sample have the following values.45,47,50,52,48,47,49,53,51. Does the mean of these values differ significantly from assumed mean 47.5? **Answer: There is no significance difference between population mean and sample mean.**	

(2). t-TEST FOR DIFFERENCE OF MEANS:

- ✓ This test is used to test whether the two samples $x_1, x_2, x_3, \dots, x_{n_1}$ and $y_1, y_2, \dots, y_{n_2}$ of sizes n_1 and $n_2$have been drawn from two normal populations with mean μ_1 and μ_2 respectively under the assumption that the population variance are equal.$(\sigma_1 = \sigma_2 = \sigma)$
- ✓ H_0 : The samples have been drawn from the normal population with means μ_1 and μ_2
- ✓ i.e., $H_0 : \mu_1 \neq \mu_2$.
- ✓ Let $\overline{X}, \overline{Y}$ be their means of the two samples.
- ✓ Under this $H_0$the test static t is given by

$$t = \frac{(\overline{X} - \overline{Y})}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- ✓ If the two sample standard deviations s_1, s_2 are given then we have

$$\sigma^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

METHOD: 7 T-TEST FOR DIFFERENCE OF MEANS

<table>
<tr>
<td>C</td>
<td>1</td>
<td>Two sample of 6 and 5 items, respectively, gave the following data.
<table>
<tr><th></th><th>1st sample</th><th>2nd sample</th></tr>
<tr><td>Mean</td><td>40</td><td>50</td></tr>
<tr><td>S.D.</td><td>8</td><td>10</td></tr>
</table>
Is the difference of the means significant?(Test at 5% level of significance)
(The value of t for 9 degree of freedom at 5% level is 2.262)
Answer: There is no significance difference between two population means.</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>2</td>
<td>Two sample of 10 and 14 items, respectively, gave the following data.
<table>
<tr><th></th><th>1st sample</th><th>2nd sample</th></tr>
<tr><td>Mean</td><td>20.3</td><td>18.6</td></tr>
<tr><td>S.D.</td><td>3.5</td><td>5.2</td></tr>
</table>
Is the difference of the means significant?(Test at 5% level of significance)
(The value of t for 22 degree of freedom at 5% level is 2.0739)
Answer: There is no significance difference between two population means.</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>A large group of teachers are trained, where some are trained by institution A and some are trained by institution B. In a random sample of 10 teachers taken from a large group, the following marks are obtained in an appropriate achievement test
<table>
<tr><td>Institution A</td><td>65</td><td>69</td><td>73</td><td>71</td><td>75</td><td>66</td><td>71</td><td>68</td><td>68</td><td>74</td></tr>
<tr><td>Institution B</td><td>78</td><td>69</td><td>72</td><td>77</td><td>84</td><td>70</td><td>73</td><td>77</td><td>75</td><td>65</td></tr>
</table>
Test the claim that institute B is more effective at 0.05 level of significance under the assumption that the two populations are normally distributed with same variances.
Answer: The claim is valid</td>
<td></td>
</tr>
</table>

<table>
<tr><td>H</td><td>4</td><td>Random samples of specimens of coal from two mines A and B are drawn and their heat producing capacity (in millions of calories per ton) were measured yielding the following result:
<table>
<tr><td>Mine a</td><td>8260</td><td>8130</td><td>8350</td><td>8070</td><td>8340</td><td></td></tr>
<tr><td>Mine B</td><td>7950</td><td>7890</td><td>7900</td><td>8140</td><td>7920</td><td>7840</td></tr>
</table>
Use the 5% level of significance to test whether the difference between the means of these two samples is significant.
Answer: The average heat producing capacity of the coal from the two mines is not the same.</td><td></td></tr>
</table>

(3). t-TEST FOR CORRELATION COEFFICIENTS:

- ✓ Consider a random sample of n observations from a bivariate normal population.Let r be the observed correlation coefficient and ρ be the population correlation coefficient.
- ✓ Under the null and alternative hypothesis as follows,
- ✓ $H_0 : \rho = 0$ (There is no correlation between two variables)
- ✓ $H_1 : \rho \neq 0$ or $\rho > 0$ or $\rho < 0$ (There is correlation between two variables)
- ✓ The test static t is given by $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$, is a t-variate with $v = n - 2$ degrees of freedom.

METHOD: 8 T-TEST FOR CORRELATION COEFFICIENTS

C	1	The correlation coefficient between income and food expenditure for sample of 7 household from a low income group is 0.9. Using 1% level of significance, test whether the correlation coefficient between incomes and food expenditure is positive. Assume that the population of both variables are normally distributed. **Answer: There is correlation between incomes and food expenditure.**	
H	2	A random sample of fifteen paired observations from a bivariate population gives a correlation coefficient of -0.5. Does this signify the existence of correlation in the sampled population?(Test at 5% level of significnace) **Answer: The sampled population is uncorrelated.**	

C	**3**	A random sample of 27 pairs of observations from a normal population gave a correlation coefficient of 0.6. Is this significant of correlation in the population? **Answer: The sampled population is correlated.**	
H	**4**	A coefficient of correlation of 0.2 is derived from a random sample of 625 pairs of observations.Is this value of r significant? **Answer: It is highly significant.**	

(4). F-TEST FOR RATIO OF VARIANCES:

- ✓ Let n_1 and n_2 be the sizes of two samples with variance s_1^2 and s_2^2.The estimate of the population variance based on these samples are $s_1 = \frac{n_1 s_1^2}{n_1 - 1}$ and$s_2 = \frac{n_2 s_2^2}{n_2 - 1}$. The degrees of freedom of these estimates are $v_1 = n_1 - 1, v_2 = n_2 - 1$.
- ✓ To test whether these estimates are significantly different or if the samples may be regarded as drawn from the same population or from two populations with same varianceσ^2. We setup the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$
- ✓ So the test static is$F = \frac{s_1^2}{s_2^2}$, the numerator is greater than denominator i.e., $s_1^2 > s_2^2$.

METHOD: 9 F-TEST FOR RATIO OF VARIANCES

<table>
<tr><td>C</td><td>1</td><td>In two independent samples of sizes 8 and 10 the sum of squares of derivations of the samples values from the respective sample means were 84.4 and 102.6. Test whether the difference of variances of the populations is significant or not.
Answer: There is no significnact difference between the variance of the populations.</td><td></td></tr>
<tr><td>H</td><td>2</td><td>The two random samples reveal the following data:
<table>
<tr><td>Sample no.</td><td>Size</td><td>Mean</td><td>Variance</td></tr>
<tr><td>I</td><td>16</td><td>440</td><td>40</td></tr>
<tr><td>II</td><td>25</td><td>460</td><td>42</td></tr>
</table>
Test whether the samples come from the same normal population.
Answer: The population variances are equal.</td><td></td></tr>
</table>

<table>
<tr><td>C</td><td>3</td><td>Two random samples drawn from 2 normal populations are as follows:
<table>
<tr><td>A</td><td>17</td><td>27</td><td>18</td><td>25</td><td>27</td><td>29</td><td>13</td><td>17</td></tr>
<tr><td>B</td><td>16</td><td>16</td><td>20</td><td>27</td><td>26</td><td>25</td><td>21</td><td></td></tr>
</table>
Test whether the samples are drawn from the same normal population.
Answer: The population variances are equal</td><td></td></tr>
<tr><td>H</td><td>4</td><td>Two independent sample of size 7 and 6 had the following values:
<table>
<tr><td>A</td><td>28</td><td>30</td><td>32</td><td>33</td><td>31</td><td>29</td><td>34</td></tr>
<tr><td>B</td><td>29</td><td>30</td><td>30</td><td>24</td><td>27</td><td>28</td><td></td></tr>
</table>
Examine whether the samples have been drawn from normal populations having the same variance.
Answer: Samples have been drawn from the normal population with same variance.</td><td></td></tr>
</table>

(5). CHI-SQUARE TEST FOR GOODNESS OF FIT

- ✓ Pare-1
  - ➢ Find the expected frequencies using general probability considerations or specific probability model (Poisson, binomial, normal) given in the problem itself.
- ✓ Part-2
  - ➢ Testing under the null and alternative hypothesis as follows.
- ✓ H_0 : Given probability distribution fits good with the given data; that is ,there is no significant difference between observed frequencies (O_i) and expected frequencies (e_i)
- ✓ H_1 : Given probability distribution does not fit good with the given data; that is , there is significant difference between observed frequencies (O_i) and expected frequencies (e_i)
- ✓ The test static given by

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \quad (with\ v = k - m\ degree\ of\ freedom)$$

METHOD: 10 CHI-SQUARE TEST FOR GOODNESS OF FIT

C **1** Suppose that a dice is tossed 120 times and the recorded data is as follows:

Face Observed(x)	1	2	3	4	5	6
Frequency	20	22	17	18	19	24

Test the hypothesis that the dice is unbiased at $\alpha = 0.05$.

Answer: The dice may be considered to be unbiased.

H **2** The following table gives the number of accidents that look place in an industry during various days of the week.Test if accidents are uniformly distributed over the week.

Day	Mon	Tue	Wed	Thus	Fri	Sat
No. of accidents	14	18	12	11	15	14

Answer: The accidents are uniformly distributed over the week.

C **3** The following table indicates (a) the frequencies of a given distribution with (b) the frequencies of the normal distribution having the same mean, standard deviation and the total frequency as in (a).

(a)	1	5	20	28	42	22	15	5	2
(b)	1	6	18	25	40	25	18	6	1

Apply the χ^2-test of goodness of fit.(take level of significance 5%).

Answer: This normal distribution provides a good fit.

H **4** Suppose that during 400 five-minute intervals the air-traffic control of an airport received $0,1,2, \dots,$ or 13 radio messages with respective frequencies of 3,15,47,76,68,74,46,39,15,9,5,2,0 and 1.Test at 0.05 level of significance, the hypothesis that the number of radio messages received during 5 minute interval follow poisson distribution with $\lambda = 4.6$.

Answer: Poisson distribution with $\boldsymbol{\lambda = 4.6}$. provides a good fit.

5 Records taken of the number of male and female births in 800 families having four children are as follows:

No. of male births	0	1	2	3	4
No. of female births	4	3	2	1	0
No. of families	32	178	290	236	94

Test whether data are consistent with hypothesis that the binomial law holds change of male birth is equal to that of female birth, namely $p = q = \frac{1}{2}$.

Answer: The data are not consistence with the hypothesis.

(6). CHI-SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES

- ✓ Pare-1
  - ➢ Construct a contingency table on the basis of given information and find expected frequency for each cell using

$$E_{ij} = \frac{\text{column total} * \text{row total}}{\text{grand total}}$$

- ✓ Part-2
  - ➢ Testing under the null and alternative hypothesis as follows.
- ✓ H_0:Attributes are independent; that is, there is no significant difference between observed frequencies (O_{ij}) and expected frequencies $\left(E_{ij}\right)$
- ✓ H_1 : Attributes are dependent; that is, there is significant difference between observed frequencies (O_{ij}) and expected frequencies $\left(E_{ij}\right)$
- ✓ The test static χ^2 for the analysis of $r \times c$ table is given by

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}} \text{ With degree of freedom } (r-1)(c-1).$$

- ✓ Here, the hypothesis $H_0$is tested using right one-tailed test.

METHOD: 11 CHI-SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES

C **1** Test the hypothesis at 0.05 level of significance that the presence or absence of hypertension is independent of smoking habits from the follo wing data of 80 persons.

	Non smokers	Moderate smokers	Heavy smokers
HT	21	36	30
No HT	48	26	19

Answer: Hypertension and smoking habits are not independent.

H **2** A company operates three machines on three different shifts daily. The following table presents the data of the machine breakdowns resulted during a 6-month time period.

Shift	Machine A	Machine B	Machine C	Total
1	12	12	11	35
2	15	25	13	53
3	17	23	10	50
Total	44	60	34	138

Test the hypothesis that for an arbitory breakdown the machine causing the breakdown and the shift on which the breakdown occurs are independent. (Take level of significance 5%).

Answer: An arbitory breakdown machine and the shift are independent.

C **3** From the following data, find whether hair color and gender are associated.

Color	Fair	Red	Medium	Dark	Black	Total
Boys	592	849	504	119	36	2100
Girls	544	677	451	97	14	1783
Total	1136	1526	955	216	50	3883

Answer: The hair color and gender are associated.

UNIT-5 » CURVE FITTING BY NUMERICAL METHOD

❖ INTRODUCTION:

- ✓ In particular statistics, we come across many situations where we often require to find a relationship between two or more variables. For example, weight and height of a person, demand and supply, expenditure depends on income, etc. These relation, in general, may be expresses by polynomial or they may have exponential or logarithmic relationship. In order to determine such relationship, first it is require to collect the data showing corresponding values of the variables under consideration.
- ✓ Suppose $(x_1, y_1), (x_2, y_2), \dots \dots, (x_n, y_n)$ be the data showing corresponding values of the variables x and y under consideration. If we plot the above data points on a rectangular coordinate system, then the set of points so plotted form a scatter diagram.
- ✓ From this diagram, it is sometimes possible to visualize a smooth curve approximating the data. Such a curve is called an approximating curve.
- ✓ In particular, if the data approximate well to a straight line, we say that a linear relationship exists between the variables. It is quite possible that the relationship of the form $y = f(x)$ between two variables x and y, giving the approximating curve and which fit the given data of x and y, is called curve fitting.

❖ CURVE FITTING:

- ✓ Curve fitting is the process of finding the 'best-fit' curve for a given set of data. It is the representation of the relationship between two variables by means of an algebraic equation.

❖ THE METHOD OF LEAST SQUARE:

- ✓ The method of least squares assumes that the best-fit curve of a given type is the curve that has the minimum sum of the square of the deviation (least square error) from a given set of data.
- ✓ Suppose that the data points are $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, where x is independent and y is dependent variable. Let the fitting curve f(x) has the following deviations (or errors or residuals) from each data points.

$$d_1 = y_1 - f(x_1), d_2 = y_2 - f(x_2), \dots , d_n = y_n - f(x_n)$$

- ✓ Clearly, some of the deviations will be positive and others negative. Thus, to give equal weightage to each error, we square each of these and form their sum; that is,

$$D = {d_1}^2 + {d_2}^2 + \cdots + {d_n}^2$$

- ✓ Now, according to the method of least squares, the best fitting curve has the property that

$$\mathbf{D = {d_1}^2 + {d_2}^2 + \cdots + {d_n}^2 = \sum_{i=1}^{n} {d_i}^2 = \sum_{i=1}^{n} [y_i - f(x_i)]^2 = a\ minimum}.$$

❖ FITTING A STRAIGHT LINE $y = a + bx$ (LINEAR APPROXIMATION):

- ✓ Suppose the equation of a straight line of the form $y = a + bx$ is to be fitted to the n-data points $(x_1, y_1), (x_2, y_2), \dots , (x_n, y_n), n \geq 2$, Where a is y-intercept and b is its slope.
- ✓ For the general point (x_i, y_i), the vertical distance of this point from the line $y = a + bx$ is the deviation d_i, then $d_i = y_i - f(x_i) = y_i - a - bx_i$.
- ✓ Applying method of least squares, the values of a and b are so determined as to minimize

$$\mathbf{D = \sum_{i=1}^{n} (y_i - a - bx_i)^2}$$

- ✓ This will be minimum, so

$$\frac{\partial D}{\partial a} = 0 \Longrightarrow -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

$$\frac{\partial D}{\partial b} = 0 \Longrightarrow -2 \sum_{i=1}^{n} x_i (y_i - a - bx_i) = 0$$

- ✓ Simplifying and expanding the above equations, we have

$$\sum_{i=1}^{n} y_i = a \sum_{i=1}^{n} 1 + b \sum_{i=1}^{n} x_i \text{ and } \sum_{i=1}^{n} x_i y_i = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} {x_i}^2$$

- ✓ Which implies

$$\sum_{i=1}^{n} y_i = an + b \sum_{i=1}^{n} x_i \text{ and } \sum_{i=1}^{n} x_i y_i = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} {x_i}^2$$

- ✓ We obtain following normal equations for the best fitting straight line $\mathbf{y = a + bx}$.

$$\sum y = an + b\sum x$$
$$\sum xy = a\sum x + b\sum x^2$$

- ✓ The normal equations for the best fitting straight line $\mathbf{y = ax + b}$ is

$$\sum y = bn + a\sum x$$
$$\sum xy = b\sum x + a\sum x^2$$

METHOD – 1: FITTING A STRAIGHT LINE

H **1** Fit a straight line for following data. Also, find y when $x = 1.8$.

X	1	2	3	4
Y	14	27	40	55

Answer: $\mathbf{y = -0.7696 + 0.782x, Y(1.8) = 0.638}$

C **2** Fit a straight line for following data. Also, find y when $x = 1.8$.

X	2	5	6	9	11
Y	2	4	6	9	10

Answer: $\mathbf{y = -0.0244 + 0.9431x, Y(2.8) = 2.6163}$

H **3** Fit a straight line for following data.

X	20.5	32.7	51.0	73.2	95.7
Y	765	826	873	942	1032

Answer: $\mathbf{y = 702.1721 + 3.3948x}$

H **4** Fit a straight line for following data.

X	0	1	2	3	4
Y	1	1.8	3.3	4.5	6.3

Answer: $\mathbf{y = 0.72 + 1.33x}$

T **5** Fit a straight line $y = ax + b$ for following data.

X	6	7	7	8	8	8	9	9	10
Y	5	5	4	5	4	3	4	3	3

Answer: $\mathbf{Y = -0.5X + 8}$

C	6	If P is the pull required to lift a load W by means of a pulley block, find a linear approximation of the form $P = mW + c$ connecting P and W, using the following data (Where P and W are taken in kg. wt.): W: 13, 18, 23, 27 P: 51, 75, 102, 119 **Answer**: $\mathbf{P = -12.95 + 4.92W}$	

W	13	18	23	27
P	51	75	102	119

❖ FITTING A PARABOLA $y = a + bx + cx^2$ BY LEAST SQUARE APPROXIMATION:

- ✓ Consider a set of n pairs of the given values (x, y) for fitting the curve $y = a + bx + cx^2$. The residual $R = y - (y = a + bx + cx^2)$ is the difference between the observed and estimated values of y. We have to find a, b, c such that the sum of the squares of the residuals is minimum (least). Let

$$\mathbf{S = \sum_{1}^{n} [y - (a + bx + cx^2)]^2} \quad \dots\dots \ (\mathbf{1})$$

- ✓ Differentiating S with respect to a, b, c and Equating zero. We obtain following normal Equations for the best fitting $y = a + bx + cx^2$ curve (parabola) of second degree.

$$\begin{aligned} &\sum \mathbf{y = n\,a + b\sum x + c\sum x^2} \\ &\sum \mathbf{x\,y = \ a\sum x + b\sum x^2 + c\sum x^3} \\ &\sum \mathbf{x^2\,y = a\sum x^2 + b\sum x^3 + c\sum x^4} \end{aligned}$$

- ✓ The normal equation for $\mathbf{y = ax^2 + bx + c}$ are

$$\begin{aligned} &\sum y = n\,c + b\sum x + a\sum x^2 \\ &\sum x\,y = \ c\sum x + b\sum x^2 + a\sum x^3 \\ &\sum x^2\,y = c\sum x^2 + b\sum x^3 + a\sum x^4 \end{aligned}$$

METHOD – 2: FITTING A PARABOLA

C **1** Fit a second degree parabola $y = a + bx + cx^2$ to the following data:

X	1.0	1.5	2.0	2.5	3.0	3.5	4.0
y=f(x)	1.2	1.4	1.9	2.4	2.8	3.3	4.2

Answer: $\mathbf{-\,0.1057 + x + 4.4286\,x^2}$

C **2** Fit a second degree parabola $y = a + bx + cx^2$ to the following data:

X	0	1	2	3	4	5	6
y	14	18	23	29	36	40	46

Answer: $\mathbf{y = 13.4523 + 4.9642x + 0.083x^2}$

H **3** Fit a second degree parabola $y = a + bx + cx^2$ to the following data:

X	0	1	2	3	4	5	6
y	14	18	23	29	36	40	46

Answer: $\mathbf{y = 13.4523 + 4.9642x + 0.083x^2}$

T **4**

X	−3	−2	−1	0	1	2	3
Y	12	4	1	2	7	15	30

Fit a second degree parabola $y = ax^2 + bx + c$ to the following data:

Answer: $\mathbf{y = 1.67x^2 + 2.93x + 2.12}$

C **5** Fit a polynomial of degree two using least square method for the following experimental data. Also estimate y(2.4).

X	1	2	3	4	5
y	5	12	26	60	97

Answer: $\mathbf{y = 10.4 + 11.09x + 5.71x^2, y(2.4) = 69.91}$

H **6** Find the least squares approximations of second degree for following data:

X	−2	−1	0	1	2
Y	15	1	1	3	19

Answer: $\mathbf{y = -0.1057 + x + 4.4286\,x^2}$

H **7** Fit a second degree parabola $y = a + bx + cx^2$ to the following data:

X	1.0	1.5	2.0	2.5	3.0	3.5	4.0
y=f(x)	1.2	1.4	1.9	2.4	2.8	3.3	4.2

Answer: $\mathbf{-0.1057 + x + 4.4286\, x^2}$

T **8** Fit a second degree parabola $y = ax^2 + bx + c$ to the following data:

X	-1	0	1	2	3
y	5	6	21	50	93

Answer: $\mathbf{y = 7x^2 + 8x + 6}$

❖ NON-POLYNOMIAL APPROXIMATION OR NON-LINEAR REGRESSION:

- ✓ $\mathbf{y = ae^{bx}}$ Yields
  - Taking Logarithm on both sides $\log y = \log a + bx$
  - Denoting $\log y = Y$ and $\log a = A$, the above equation becomes $Y = A + bx$
  - Find A, b & consequently a=Antilog A can be calculated.
- ✓ $\mathbf{y = ax^b}$ **Yields**
  - Taking Logarithm on both sides $\log y = \log a + b \log x$
  - Denoting $\log y = Y$, $\log a = A$ and $\log x = X$, we obtain$Y = A + bX$
  - Find A, b can be found & consequently a=Antilog A can be calculated.
- ✓ $\mathbf{y = ab^x}$ **Yields**
  - Taking Logarithm on both sides $\log y = \log a + x \log b$
  - Denoting $\log y = Y$, $\log a = A$, and $\log b = B$, we obtain $Y = A + Bx$
  - Find A, B can be found & consequently a=Antilog A and b=Antilog B can be calculated.

METHOD – 3: FITTING THE GENERAL CURVES

C **1** Determine a and b so that $y = ae^{bx}$ fits the data :

X	1	2	3	4
Y	7	11	17	27

Answer: $\mathbf{a = 4.4680, b = 0.4485}$

H **2** Fit a curve of the form $y = ab^x$ for the data and hence find the estimation for y when $x = 8$.

X	1	2	3	4	5	6	7
Y	87	97	113	129	202	195	193

Answer: $\mathbf{a = 73.7416, b = 0.1560, y(8) = 256.8315}$

H **3** Fit a curve of the form $y = ax^b$ to the following data.

X	20	16	10	11	14
Y	22	41	120	89	56

Answer: $\mathbf{a = 27181.4811, b = -2.3608}$

H **4** By method of least square fit a curve of the form $y = ax^b$ to the following data:

X	2	3	4	5
y	27.8	62.1	110	161

Answer: $\mathbf{a = 7.3883, b = 1.9293}$

C **5** Find the least square fit of the form $y = a_0 + a_1x^2$ to the following data:

x	-1	0	1	2
Y	2	5	3	0

Answer: $\mathbf{y = 4.167 - 1.111x^2}$

C **6** Using least square method fit the curve $y = ax^2 + \frac{b}{x}$ to the following data:

x:	1	2	3	4
y:	-1.51	0.99	8.88	7.66

Answer: $\mathbf{y = 0.51x^2 - 2.06/x}$

GUJARAT TECHNOLOGICAL UNIVERSITY

Bachelor of Engineering
Subject Code: 3130006
Semester – III
Subject Name: Probability and Statistics

Type of course: Basic Science Course

Prerequisite: Probability basics

Rationale: Systematic study of uncertainty (randomness) by probability - statistics and curve fitting by numerical methods

Teaching and Examination Scheme:

Teaching Scheme			Credits	Examination Marks				Total Marks
L	T	P	C	Theory Marks		Practical Marks		
				ESE (E)	PA (M)	ESE (V)	PA (I)	
3	2	0	5	70	30	0	0	100

Content:

Sr. No.	**Content**	**Total Hrs**	**% Weightage**
01	**Basic Probability:** Experiment, definition of probability, conditional probability, independent events, Bayes' rule, Bernoulli trials, Random variables, discrete random variable, probability mass function, continuous random variable, probability density function, cumulative distribution function, properties of cumulative distribution function, Two dimensional random variables and their distribution functions, Marginal probability function, Independent random variables.	08	20 %
02	**Some special Probability Distributions**: Binomial distribution, Poisson distribution, Poisson approximation to the binomial distribution, Normal, Exponential and Gamma densities, Evaluation of statistical parameters for these distributions.	10	25 %
03	**Basic Statistics:** Measure of central tendency: Moments, Expectation, dispersion, skewness, kurtosis, expected value of two dimensional random variable, Linear Correlation, correlation coefficient, rank correlation coefficient, Regression, Bounds on probability, Chebyshev‘s Inequality	10	20%
04	**Applied Statistics:** Formation of Hypothesis, Test of significance: Large sample test for single proportion, Difference of proportions, Single mean, Difference of means, and Difference of standard deviations. Test of significance for Small samples: t- Test for single mean, difference of means, t-test for correlation coefficients, F- test for ratio of variances, Chi-square test for goodness of fit and independence of attributes.	10	25 %
05	Curve fitting by the numerical method: Curve fitting by of method of least squares, fitting of straight lines, second degree parabola and more general curves.	04	10 %

Suggested Specification table with Marks (Theory):

Distribution of Theory Marks					
R Level	U Level	A Level	N Level	E Level	C Level
7	**28**	**35**	0	0	0

Legends: R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create and above Levels (Revised Bloom's Taxonomy)

Note: This specification table shall be treated as a general guideline for students and teachers. The actual distribution of marks in the question paper may vary from above table. This subject will be taught by Maths faculties.

Reference Books:

(1) P. G. Hoel, S. C. Port and C. J. Stone, Introduction to Probability Theory, Universal Book Stall.
(2) S. Ross, A First Course in Probability, 6th Ed., Pearson Education India.
(3) W. Feller, An Introduction to Probability Theory and its Applications, Vol. 1, Wiley.
(4) D. C. Montgomery and G. C. Runger, Applied Statistics and Probability for Engineers, Wiley.
(5) J. L. Devore, Probability and Statistics for Engineering and the Sciences, Cengage Learning.

Course Outcome:

Sr. No.	CO statement	Marks % weightage
CO-1	understand the terminologies of basic probability, two types of random variables and their probability functions	20 %
CO-2	observe and analyze the behavior of various discrete and continuous probability distributions	25 %
CO-3	understand the central tendency, correlation and correlation coefficient and also regression	20%
CO-4	apply the statistics for testing the significance of the given large and small sample data by using t- test, F- test and Chi-square test	25 %
CO-5	understand the fitting of various curves by method of least square	10 %

List of Open Source Software/learning website:

MIT Opencourseware. NPTEL.