Dedicated Faculty, Committed Education
Darshan
Institute of Engineering & Technology

3rd Edition

Probability & Statistics

(3130006)

Computer Engineering

Name : - ____________________

Roll No. : - ____________________

Division : - ____________________

DARSHAN INSTITUTE OF ENGINEERING & TECHNOLOGY

INDEX

INDEX

UNIT 1 – BASIC PROBABILITY THEORY

❖ 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 EXPERIMENT

- ✓ Random experiment is an experiment about whom outcomes cannot be successfully predicted. Of course, we know all possible outcomes in advance.

❖ SAMPLE SPACE

- ✓ The set of all possible outcomes of a random experiment is called a sample space.
- ✓ It is denoted by “S” and 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.

$$S = \{H, T\} \times \{H, T\} = \{HH, HT, TH, TT\}$$

- ✓ Infinite Sample Space: Experiment of tossing a coin until a head comes up for first time.

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

❖ EVENT

- ✓ A subset of a sample space is known as Event. Each member is called Sample Point.
- ✓ Example:
  - ➢ Experiment: Tossing a coin twice. $S = \{HH, HT, TH, TT\}$

  - Event A: Getting TAIL both times. $A = \{TT\}$
  - Event B: Getting TAIL exactly once. $B = \{HT, TH\}$

❖ DEFINITIONS

- ✓ The subset $\emptyset$ 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".
- ✓ A 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: $A \cap B = \{ x \mid x \in A \text{ AND } x \in B \}$
- ✓ If $A \cup B = S$. Events are called Mutually Exhaustive Events.
  - Set Notation: $A \cup B = \{ x \mid x \in A \text{ OR } x \in B \}$
- ✓ If $A \cap B = \phi$ and $A \cup B = S$. Events are called Mutually Exclusive & Exhaustive Events.

METHOD – 1: BASIC EXAMPLES ON SAMPLE SPACE AND EVENT

H	1	Define Mutually Exclusive and Exhaustive events with a suitable example.	
C	2	A coin is tossed twice, and their up faces are recorded. What is the sample space for this experiment? **Answer**: $\mathbf{S = \{HH, HT, TH, TT\}}$	
C	3	Suppose a pair of dice are tossed. What is the sample space for the experiment? **Answer**: $\mathbf{S = \{(1,1), (1,2), \ldots (1,6), \ldots \ldots \ldots, (6,1), (6,2), \ldots (6,6)\}}$	
H	4	Four cards are labeled with A, B, C and D. We select two cards at random without replacement. Describe the sample space for the experiments. **Answer**: $\mathbf{S = \{AB, AC, AD, BC, BD, CD\}}$	

H	5	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**: $\mathbf{S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}}$ $\mathbf{S = \{H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6\}}$	
H	6	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**: $\mathbf{S = \{\, HHH, HHT, HTH, HTT, THH, THT, TTH, TTT1, TTT2,}$ $\mathbf{TTT3, TTT4, TTT5, TTT6\}}$	
C	7	Two unbiased dice are thrown. Write down the following events: Event A: Both the dice show the same number. Event B: The total of the numbers on the dice is 8. Event C: The total of the numbers on the dice is 13. Event D: The total of the number on the dice is any number from $[2,\ 12]$. **Answer**: $\mathbf{A = \{\, (1,\ 1),\ (2,\ 2),\ (3,\ 3),\ (4,\ 4),\ (5,\ 5),\ (6,\ 6)\, \}}$ $\mathbf{B = \{\, (2,\ 6),\ (3,\ 5),\ (4,\ 4),\ (5,\ 3),\ (6,\ 2)}$ $\mathbf{C = \{\emptyset\}}$ $\mathbf{D = \{\, (1,\ 1),\ ..,\ (1,\ 6), \dots,\ (6,\ 1),\ ..,\ (6,\ 6)\, \}}$	
H	8	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**: $\mathbf{S = \{HR_1, HR_2, HR_3, HG_1, HG_2, HG_3, HG_4, T1, T2, T3, T4, T5, T6\}}$	
H	9	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 **Answer**: $\mathbf{S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}}$ $\mathbf{EVENT\ A = \{HHH, HHT, HTH, THH\}}$ $\mathbf{EVENT\ B = \{HTT, THT, TTH\}}$ $\mathbf{EVENT\ C = \{HHH, HHT, HTH, THH\}}$ $\mathbf{EVENT\ D = \{HHT, HTH, HTT, THH, THT, TTH, TTT\}}$	

PROBABILITY OF AN EVENT

- ✓ If a finite sample space associated with a random experiment has "n" equally likely (Equiprobable) outcomes (elements) and of these "m“ $(0 \leq m \leq n)$ outcomes are favorable for the occurrence of an event A, then probability of A is defined as below.

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

EQUIPROBABLE EVENTS

- ✓ Let $U = \{x_1, x_2, \dots, x_n\}$ 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 $A \subset B$, then $P(B - A) = P(B) - P(A)$ and $P(A) \leq P(B)$.
- ✓ For every event A, $0 \leq P(A) \leq 1$.
- ✓ Let S be sample space and A, B and C be any events in S, then
  - ➢ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - ➢ $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$
  - ➢ $P(A \cap B') = P(A) - P(A \cap B)$
  - ➢ $P(A' \cap B) = P(B) - P(A \cap B)$
  - ➢ $P(A' \cap B') = P(A \cup B)' = 1 - P(A \cup B)$ (De Morgan’s Rule)
  - ➢ $P(A' \cup B') = P(A \cap B)' = 1 - P(A \cap B)$ (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 rth object, it follows by the fundamental principle of counting that the number of different arrangement (or PERMUTATIONS) is given as below.

$$^{n}P_{r} = n\,(n-1)\,(n-2)\,\dots\,(n-r+1) = \frac{n!}{(n-r)!}$$

❖ RESULTS ON PERMUTATION

- ✓ 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

$$\frac{n!}{n_1!\ n_2!\ \dots\ n_k!}.$$

  - ➢ A number of different permutations of letters of the word MISSISSIPPI is

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

- ✓ If 'r' objects are to be arranged out of 'n' objects and if repetition of an object is allowed then the total number of permutations is n^r.

  - ➢ 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

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

❖ EXAMPLES ON COMBINATION

- ✓ The number of ways in which 3 card can be chosen from 8 cards is

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

- ✓ 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?

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

- ✓ 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 included?

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

METHOD – 2: EXAMPLES ON PROBABILITY OF EVENTS

C	1	If probability of event A is $\frac{9}{10}$, what is the probability of the event "not A"? **Answer**: $\mathbf{0.1}$	
C	2	What is the probability that a leap year contains 53 Sundays? **Answer**: $\mathbf{0.2857}$	
H	3	Three coins are tossed. Find the probability of (a) Getting at least 2 heads, (b) Getting exactly 2 head. **Answer**: $\mathbf{0.5,\ \ 0.375}$	
H	4	A single die is tossed once. Find the probability of a 2 or 5 turning up. **Answer**: $\mathbf{\frac{1}{3}}$	
H	5	Two unbiased dice are thrown. Find the probability that: (a) Both the dice show the same number. (b) The first die shows 6. (c) The total of the numbers on the dice is 8. (d) The total of the numbers on the dice is greater than 8. (e) The total of the numbers on the dice is 13. (f) Total of numbers on the dice is any number from 2 to 12, both inclusive. **Answer**: $\mathbf{\frac{1}{6},\ \frac{1}{6},\ \frac{5}{36},\ \frac{5}{18},\ \ 0,\ \ 1}$	
C	6	(a) A club has 5 male and 7 female members. A committee composed of 3 men and 4 women is formed. In how many ways this be done? (b) 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 included? **Answer**: $\mathbf{350,\ \ 186}$	

H	7	One card is drawn at random from a well shuffled pack of 52 cards. Find probability that the card will be (a) an ace, (b) a card of black color, (c) a diamond, (d) not an ace. **Answer**: $\mathbf{0.0769,\ \ 0.5,\ \ 0.25,\ \ 0.9231}$	
C	8	If 5 cards are drawn from a pack of 52 well-shuffled cards, find the probability of (a) 4 ace, (b) 4 aces and 1 is a king, (c) 3 are tens and 2 are jacks, (d) a nine, ten, jack, queen, king is obtained in any order, (e) 3 are of any one suit and 2 are of another, (f) at least one ace is obtained. **Answer**: $\mathbf{\frac{1}{54145},\ \frac{1}{649740},\ \frac{1}{108290},\ \frac{64}{162435},\ \frac{429}{4165},\ \frac{18472}{54145}}$	
H	9	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. **Answer**: $\mathbf{0.0026,\ \ 0.1055,\ \ 0.0225}$	
T	10	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**: $\mathbf{\frac{1}{4165}}$	
H	11	4 cards are drawn at random from a pack of 52 cards. Find probability that (a) They are a king, a queen, a jack and an ace. (b) Two are kings and two are queens. (c) Two are black and two are red. (d) There are two cards of hearts and two cards of diamonds. **Answer**: $\mathbf{0.00095,\ \ 0.00013,\ \ 0.3902,\ \ 0.0225}$	

H	12	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**: $\frac{2}{13}, \ \frac{5}{13}$	
H	13	There are 5 yellow, 2 red, and 3 white balls in the box. Three balls are randomly selected from the box. Find the probability of the following events. (a) all are of different color, (b) 2 yellow and 1 red color, (c) all are of same color. **Answer**: $0.25, \ 0.1667, \ 0.0917$	
C	14	An urn contains 6 green, 4 red and 9 black balls. If 3 balls are drawn at random, find the probability that at least one is green. **Answer**: $\frac{683}{969}$	
T	15	A box contains 6 red balls, 4 white balls, 5 black balls. A person draws 4 balls from the box at random. Find the probability that among the balls drawn there is at least one ball of each color. **Answer**: 0.5275	
T	16	A machine produces a total of 12000 bolts a day, which are on the average 3% defective. Find the probability that out 600 bolts chosen at random, 12 will be defective. **Answer**: $\frac{\binom{360}{12}\binom{11640}{588}}{\binom{12000}{600}}$	
C	17	If 5 of 20 tires in storage are defective and 5 of them are randomly chosen for inspection (that is, each tire has the same chance of being selected), what is the probability that the two of the defective tires will be included? **Answer**: 0.2935	W-19 (4)

C	18	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**: $\frac{29}{30}$	
H	19	Do as directed: (a) Find the probability that there will be 5 Sundays in the month of July. (b) Find the probability that there will be 5 Sundays in the month of June. (c) What is the probability that a non-leap year contains 53 Sundays? (d) What is the probability that a leap year contains 53 Sundays? **Answer**: $\frac{3}{7}, \frac{2}{7}, \frac{1}{7}, \frac{2}{7}$	
H	20	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**: $\mathbf{0.7}$, $\mathbf{0}$, $\mathbf{0.75}$, $\mathbf{0.25}$	
C	21	The probability that a student passes a physics test is $\frac{2}{3}$ and the probability that he passes both physics and English tests is $\frac{14}{45}$. The probability that he passes at least one test is $\frac{4}{5}$, what is the probability that he passes the English test? **Answer**: $\frac{4}{9}$	
H	22	A basket contains 20 apples and 10 oranges of which 5 apples and 3 oranges are bad. If a person takes 2 at random, what is the probability that either both are apples or both are good? **Answer**: $\frac{316}{435}$	
C	23	Two dice are thrown together. Find the probability that the sum is divisible by 2 or 3. **Answer**: $\mathbf{0.6667}$	

H	24	A card is drawn from a pack of 52 cards. Find the probability of getting a king or a heart or a red card. **Answer**: $\frac{7}{13}$	
H	25	An integer is chosen at random from the first 200 positive integers. What is the probability that the integer is divisible by 6 or 8? **Answer**: 0.25	
T	26	Three newspapers A, B, C are published in a certain city. It is estimated from a survey that of the adult population: 20% read A, 16% read B, 14% read C, 8% read both A and B, 5% read both A and C, 4% read B and C, 2% read all three. Find what percentage read at least one of the papers? **Answer**: 35%	
T	27	Four letters of the word THURSDAY are arranged in all possible ways. Find the probability that the word formed is HURT. **Answer**: $\frac{1}{1680}$	
H	28	A class has 10 boys and 5 girls. Three students are selected at random one after the other. Find the probability that (a) First two are boys and third is girl. (b) First and third of same gender and second is of opposite gender. **Answer**: $\frac{15}{91}, \frac{5}{21}$	
H	29	In how many different ways can 4 of 15 laboratory assistants be chosen to assist with an experiment? **Answer**: 1365	W-19 (3)

<table>
<tr><td>H</td><td>30</td><td>A market survey was conducted in four cities to find out the preference for brand A soap. The responses are shown below:
<table>
<tr><th></th><th>Delhi</th><th>Kolkata</th><th>Chennai</th><th>Mumbai</th></tr>
<tr><td>Yes</td><td>45</td><td>55</td><td>60</td><td>50</td></tr>
<tr><td>No</td><td>35</td><td>45</td><td>35</td><td>45</td></tr>
<tr><td>No opinion</td><td>5</td><td>5</td><td>5</td><td>5</td></tr>
</table>
(a) What is the probability that a consumer preferred brand A, given that he was from Chennai?
(b) Given that a consumer preferred brand A, what is the probability that he was from Mumbai?
Answer: $\mathbf{0.6}$, $\mathbf{0.5}$</td><td></td></tr>
<tr><td>H</td><td>31</td><td>If 3 balls are “randomly drawn” from a bowl containing 6 white and 5 black balls. What is the probability that one of the balls is white and the other two black?
Answer: $\mathbf{0.3636}$</td><td>W-19
(3)</td></tr>
<tr><td>T</td><td>32</td><td>A card from a pack of 52 cards is lost. From the remaining cards of pack, two cards are drawn and are found to be hearts. Find the probability of the missing card to be a heart.
Answer: $\mathbf{\frac{11}{50}}$</td><td></td></tr>
</table>

❖ 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

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

- ✓ 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
  - $P(A \cap B) = P(A) \cdot P(B)$ and $P(B \cap C) = P(B) \cdot P(C)$.
  - $P(C \cap A) = P(C) \cdot P(A)$ and $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$.
- ✓ Let A, B and C are said to be Pairwise independent, if
  - $P(A \cap B) = P(A) \cdot P(B), P(B \cap C) = P(B) \cdot P(C)$ & $P(C \cap A) = P(C) \cdot P(A)$.

METHOD – 3: EXAMPLES ON CONDITIONAL PROBABILITY

C	1	If $P(A) = \frac{1}{3}, P(B) = \frac{3}{4}$ and $P(A \cup B) = \frac{11}{12}$. Find $P(A/B)$. **Answer:** $\frac{2}{9}$	

H	2	If $P(A)=\frac{1}{3}, P(B)=\frac{1}{4}, P(A\cup B)=\frac{1}{2}$, then find $P(B/A), P(A/B')$. **Answer**: $\frac{1}{4}, \frac{1}{3}$	
C	3	$P(A)=\frac{1}{3}, P(B')=\frac{1}{4}, P(A\cap B)=\frac{1}{6}$, then find $P(A\cup B)$, $P(A'\cap B')$ and $P(A'/B')$. **Answer**: $\frac{11}{12}, \frac{1}{12}, \frac{1}{3}$	
C	4	A card is drawn from a well-shuffled deck of 52 cards and then second card is drawn, find the probability that one card is a spade and then second card is club if the first card is not replaced. **Answer**: $\frac{13}{204}$	
H	5	In a group of 200 students 40 are taking English, 50 are taking math, 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 math. 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 math? **Answer**: 0.20, 0.24, 0.3	
H	6	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, (b) a bulb to be drawn at random has no B type defect under the condition that it has no A type defect. **Answer**: 0.2, 0.9667	
C	7	In a certain college 25% of the students failed in probability and 15% of the student failed in statistics. A student is selected at random and 10% of the students failed in both. If he failed in probability, what is probability that he failed in statistics? **Answer**: 0.4	

H	8	Two integers are selected at random from 1 to 11. If the sum is even, find the probability that both the integers are odd. **Answer**: $\mathbf{0.6}$	
C	9	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 one is white and one is black. **Answer**: $\mathbf{\frac{4}{15}}$	
C	10	In producing screws, let A mean “screw too slim” and B “screw too short”. Let $p(A)=0.1$ and $P(B\cap A)=0.02$. A screw, selected randomly, is of type A, what is probability that a screw is of type B. **Answer**: $\mathbf{0.2}$	
H	11	A bag contains 6 white, 9 black balls. 4 balls are drawn at a time. Find the probability for first draw to give 4 white & second draw to give 4 black balls in each of following cases. (a) The balls are replaced before the second draw. (b) The balls are not replaced before the second draw. **Answer**: $\mathbf{\frac{6}{5915}, \frac{3}{715}}$	
H	12	For two independent events A & B if $P(A)=0.3,\ P(A\cup B)=0.6$, find $P(B)$. **Answer**: $\mathbf{0.4286}$	
H	13	If A, B are independent events and $P(A)=1/4, P(B)=2/3$. Find $P(A\cup B)$. **Answer**: $\mathbf{0.75}$	
C	14	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**: $\mathbf{\frac{59}{64}, \frac{3}{8}, \frac{7}{8}}$	
C	15	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	16	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 probability that target will be hit? **Answer**: $\frac{\mathbf{11}}{\mathbf{12}}$	
T	17	A problem in statistics is given to three students A, B, C whose chances of solving it are 0.5, 0.75 and 0.25 respectively. What is the probability that the problem will be solved if all of them try independently? **Answer**: $\frac{\mathbf{29}}{\mathbf{32}}$	
T	18	If A and B are independent events with $P(A) = 0.26$, $P(B) = 0.45$, find (a) $P(A \cap B)$; (b) $P(A \cap \overline{B})$; (c) $P(\overline{A} \cap \overline{B})$. **Answer**: $\mathbf{0.117}$, $\mathbf{0.143}$, $\mathbf{0.407}$	
H	19	Show that A and B are independent events if $P(A) = 0.20$, $P(B) = 0.40$ and $P(A \cup B) = 0.50$.	

❖ TOTAL PROBABILITY

- ✓ If B_1 & 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,

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

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

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

METHOD – 4: EXAMPLES ON TOTAL PROBABILITY AND BAYES' THEOREM

C	1	Consider two boxes, first with 5-green & 2-pink and second with 4-green & 3-pink balls. Two balls are selected from random box. If both balls are pink, find the probability that they are from second box. **Answer**: $\frac{3}{4}$	
H	2	In a certain assembly plant, three machines, B_1, B_2 and B_3, 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**: $\mathbf{0.0245}$	
H	3	There are three boxes. Box I contains 10 light bulbs of which 4 are defective. Box II contains 6 light bulbs of which 1 is defective and box III contains 8 light bulbs of which 3 are defective. A box is chosen and a bulb is drawn. Find the probability that the bulb is defective. **Answer**: $\mathbf{0.3139}$	
T	4	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**: $\frac{59}{130}$	
C	5	Suppose that the population of a certain city is 40% male & 60% female. Suppose also that 50% of males & 30% of females smoke. Find the probability that a smoker is male. **Answer**: $\frac{10}{19}$	

H	6	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**: $\mathbf{0.3824}$	W-19 (4)
H	7	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**: $\mathbf{0.3623}$, $\mathbf{0.4058}$, $\mathbf{0.2319}$	
H	8	A company has two plants to manufacture hydraulic machine. Plant 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**: $\mathbf{0.6747}$	
H	9	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**: $\mathbf{\frac{40}{61}}$	
H	10	Urn A contain 1 white, 2 black, 3 red balls; Urn B contain 2 white, 1 black, 1 red balls; Urn C contain 4 white, 5 black, 3 red balls. One urn is chosen at random & two balls are drawn. These happen to be one white & one red. What is probability that they come from urn A? **Answer**: $\mathbf{0.2797}$	

C	11	Three hospitals contain 10%, 20% and 30% of diabetes patients. A Patient is selected at random who is diabetes patient. Determine the probability that this patient comes from first hospital. **Answer**: $\mathbf{0.1667}$	
C	12	In a computer engineering class, 5% of the boys and 10% of the girls have an IQ of more than 150. In this class, 60% of student are boys. If a student is selected random and found to have IQ more than 150, find the probability that the student is a boy. **Answer**: $\mathbf{\frac{3}{7}}$	
H	13	A factory has three machines X, Y, Z producing $1000, 2000, 3000$ bolts per day respectively. Machine X produces 1% defective bolts, Y produces 1.5%, Z produces 2% defective bolts. At end of the day, a bolt is drawn at random and it is found to be defective. What is the probability that this defective bolt has been produced by the machine X? **Answer**: $\mathbf{0.1}$	
T	14	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**: $\mathbf{\frac{2}{3}}$	
T	15	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**: $\mathbf{0.002}$	

C	16	An insurance company insured 2000 bike drivers, 4000 car drivers and 6000 truck drivers. The probability of an accident involving a bike driver, a car driver and a truck driver is $0.10, 0.03$ and 0.15 respectively. One of the insured persons meets with an accident. What is the probability that he is a bike driver? **Answer**: $\mathbf{0.1639}$	

❖ RANDOM VARIABLE

- ✓ A random variable is a variable whose value is unknown or a function that assigns values to each of an experiment's outcomes. Random variables are often designated by capital letters X, Y.
- ✓ Random variables can be classified as
  1) Discrete Random variables, which are variable that have specific values.
  2) Continuous Random variables, which are variables that can have any values within a continuous range.

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

$$\text{where } p(x_i) \geq 0 \text{ and } \sum_i p(x_i) = 1 \text{ for all i.}$$

- ✓ **Example:** Two balanced coins are tossed, find the probability distribution for heads.
  - ➢ Sample space $= \{HH, HT, TH, TT\}$.
  - ➢ $P(X = 0) = P(\text{no head}) = \frac{1}{4} = 0.25$ and $P(X = 1) = P(\text{one head}) = \frac{2}{4} = \frac{1}{2} = 0.5$.
  - ➢ $P(X = 2) = P(\text{two heads}) = \frac{1}{4} = 0.25$.
  - ➢ Probability distribution is as follow:

X	0	1	2
P(X=x)	0.25	0.5	0.25

❖ DISCRETE RANDOM VARIABLE

- ✓ A random variable, which can take only finite, countable, or isolated values in a given interval, is called discrete random variable.
- ✓ 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 2 coins.
- ✓ 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.
- ✓ 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.
- ✓ 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 functions can be classified as (1) Probability Mass Function (P. M. F.) or (2) Probability Density Function (P. D. F.).

❖ 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.
- ✓ Conditions:

- $p(x_i) \geq 0$ for all i.
- $\sum_{i=1}^{n} p(x_i) = 1$.

❖ 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.
- ✓ Conditions:
    - $f(x_i) \geq 0$ for all i.
    - $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
    - $P(a < x < b) = \int_{a}^{b} f(x)\, dx$.

❖ MATHEMATICAL EXPECTATION

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

$$E(X) = \sum_{i=1}^{n} x_i \cdot P(x_i).$$

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

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

- ✓ 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. $E(c) = c$.
    - If c is constant, then $E(cX) = c \cdot E(X)$.
    - $E(X^2) = \sum_{i=1}^{n} x_i^2 \cdot P(x_i)$ (PMF).
    - $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx$ (PDF).
    - If a and b are constants, then $E(aX \pm b) = aE(X) \pm b$.
    - If a , b and c are constants, then $E\left(\frac{aX+b}{c}\right) = \frac{1}{c}[aE(X) + b]$.

- If X and Y are two random variables, then $E(X+Y)=E(X)+E(Y)$.
- If X and Y are two independent random variables, then $E(X\cdot Y)=E(X)\cdot E(Y)$.

❖ 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 $P(X)$ (or probability density function), then expected value of $[X-E(X)]^2$ is called the variance of X and it is denoted by $V(X)$.

$$V(X)=E(X^2)-[E(X)]^2$$

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

❖ 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., $\sigma=\sqrt{V(X)}$.
- ✓ σ^2 is called variance of $V(X)$.

❖ DISCRETE DISTRIBUTION FUNCTION

- ✓ Let X be a discrete random variable which takes the values $x_1, x_2, \ldots$ such that $x_1<x_2<\cdots$ with probabilities $P(x_1), P(x_2), \ldots$ such that $P(x_i)\geq 0$ for all values of i and

$$\sum_{i=1}^{x} P(x_i)=1.$$

- ✓ The distribution function $F(x)$ of the discrete random variable X is defined by

$$F(x)=P(X\leq x)=\sum_{i=1}^{x} P(x_i).$$

- ✓ Where x is any integer. The function $F(x)$ is also called the cumulative distribution function. The set of pairs $\{x_i, F(x)\}, i = 1,2, \dots$ is called the cumulative probability distribution.

X	x_1	x_2	…
$F(x)$	$P(x_1)$	$P(x_1) + P(x_2)$	…

❖ CONTINUOUS DISTRIBUTION FUNCTION

- ✓ If X is a continuous random variable having the probability density function $f(x)$ then the function

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx, -\infty < x < \infty.$$

is called the distribution function OR cumulative distribution function of the random variable X.

- ✓ Properties of Cumulative Distribution Function:
  - ➢ $F(-\infty) = 0, F(+\infty) = 1$ and $0 \le F(x) \le 1, \quad -\infty < x < \infty.$
  - ➢ $P(\{x_1 < X < x_2\}) = F(x_2) - F(x_1).$
  - ➢ $P(X > x) = 1 - F(x).$
  - ➢ $F'(x) = \frac{d}{dx} F(x) = f(x), \quad f(x) \ge 0.$
  - ➢ 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 $x \le x_0$.

METHOD – 5: EXAMPLES ON RANDOM VARIABLE

C	1	Which of the following functions are probability function? (a) $P(X = x) = \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{1-x}; x = 0,1$ (b) $P(X = x) = \left(-\frac{1}{2}\right)^x; x = 0,1,2$ **Answer: yes, no**	

<table>
<tr><td>H</td><td>2</td><td>Find expected value of a random variable X having following probability distribution.
<table>
<tr><td>X</td><td>-5</td><td>-1</td><td>0</td><td>1</td><td>5</td><td>8</td></tr>
<tr><td>$P(X = x)$</td><td>0.12</td><td>0.16</td><td>0.28</td><td>0.22</td><td>0.12</td><td>0.1</td></tr>
</table>
Answer: $\mathbf{0.86}$</td><td></td></tr>
<tr><td>C</td><td>3</td><td>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.
<table>
<tr><td>Number of malfunctions x</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr>
<tr><td>Probability f(x)</td><td>0.17</td><td>0.29</td><td>0.27</td><td>0.16</td><td>0.07</td><td>0.03</td><td>0.01</td></tr>
</table>
Find the mean and variance of this probability distribution.
Answer: $\mathbf{1.8}$, $\mathbf{1.8}$</td><td></td></tr>
<tr><td>H</td><td>4</td><td>The probability distribution of a random variable x is as follows. Find p and E(x).
<table>
<tr><td>X</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>P(x)</td><td>p</td><td>$\frac{1}{5}$</td><td>$\frac{1}{10}$</td><td>p</td><td>$\frac{1}{20}$</td><td>$\frac{1}{20}$</td></tr>
</table>
Answer: $\mathbf{P = 0.30}$, $\mathbf{E(x) = 1.7500}$</td><td></td></tr>
<tr><td>C</td><td>5</td><td>A random variable X has the following function.
<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></tr>
<tr><td>$P(X = x)$</td><td>0</td><td>K</td><td>2k</td><td>2K</td><td>3k</td><td>k^2</td><td>$2k^2$</td><td>$7k^2 + k$</td></tr>
</table>
Find the value of k and then evaluate $P(X < 6), P(X \geq 6)$ and $P(0 < x < 5)$.
Answer: $\mathbf{0.1}$, $\mathbf{0.81}$, $\mathbf{0.19}$, $\mathbf{0.8}$</td><td></td></tr>
<tr><td>C</td><td>6</td><td>Probability distribution of a random variable X is given below. Find $E(X), V(X), \sigma(X), E(3X + 2), V(3X + 2)$.
<table>
<tr><td>X</td><td>1</td><td>2</td><td>3</td><td>4</td></tr>
<tr><td>P(X)</td><td>0.1</td><td>0.2</td><td>0.5</td><td>0.2</td></tr>
</table>
Answer: $\mathbf{2.8}$, $\mathbf{0.76}$, $\mathbf{0.8718}$, $\mathbf{10.4}$, $\mathbf{6.84}$</td><td></td></tr>
</table>

<table>
<tr><td>H</td><td>7</td><td>Probability distribution of a random variable X is given below. Find $P(2 \le x \le 4)$ and $P(x > 2)$.
<table>
<tr><td>X</td><td>1</td><td>2</td><td>3</td><td>4</td></tr>
<tr><td>P(X)</td><td>0.1</td><td>0.2</td><td>0.5</td><td>0.2</td></tr>
</table>
Answer: $\mathbf{0.9,\ \ 0.7}$</td><td></td></tr>
<tr><td>H</td><td>8</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>$\frac{1}{12}$</td><td>$\frac{1}{3}$</td><td>a</td><td>$\frac{1}{4}$</td><td>$\frac{1}{6}$</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>
<tr><td>H</td><td>9</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>$\frac{3}{10}$</td><td>$\frac{1}{10}$</td><td>k</td><td>$\frac{3}{10}$</td><td>$\frac{1}{10}$</td></tr>
</table>
Find $k, E(X), E\,(4X+3), E(X^2), V(X), V(2X+3)$.
Answer: $\mathbf{\frac{1}{5}, \frac{4}{5}, \frac{31}{5}, \frac{13}{5}, \frac{49}{25}, \frac{196}{25}}$</td><td></td></tr>
<tr><td>C</td><td>10</td><td>If $P(x) = \frac{2x+1}{48}, x = 1, 2, 3, 4, 5, 6,$ verify whether $p(x)$ is probability function.
Answer: yes</td><td></td></tr>
<tr><td>H</td><td>11</td><td>If $P(X = x) = \frac{x}{15}, x = 1$ to 5. Find P(1 or 2) & $P(\{0.5 < X < 2.5\}/\{X > 1\})$.
Answer: $\mathbf{0.1911,\ \ 0.1429}$</td><td></td></tr>
<tr><td>C</td><td>12</td><td>Find ‘k’ for the probability distribution $p(x) = k\binom{4}{x}, x = 0, 1, 2, 3, 4.$
Answer: $\mathbf{\frac{1}{16}}$</td><td></td></tr>
<tr><td>T</td><td>13</td><td>Let mean and standard deviation of a random variable X be 5 & 5 respectively, find $E(X^2)$ and $E(2X+5)^2$.
Answer: $\mathbf{50,\ \ 325}$</td><td></td></tr>
</table>

C	14	Three balanced coins are tossed, find the mathematical expectation of tails. **Answer**: $\mathbf{1.5}$	
T	15	4 raw mangoes are mixed accidentally with the 16 ripe mangoes. Find the probability distribution of the raw mangoes in a draw of 2 mangoes. **Answer**: $\frac{60}{95}, \frac{32}{95}, \frac{3}{95}$	
C	16	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**: $\mathbf{468}$	
H	17	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**: $\mathbf{800}$	
C	18	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**: $\mathbf{0.75}$	
C	19	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**: $\mathbf{32}$	
H	20	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**: $\mathbf{1.2}$	
C	21	(a) A contestant tosses a coin and receives \$5 if head 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**: $\mathbf{\$3.00}$, $\mathbf{\$0.50}$	

C	22	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**: $\frac{1}{9}, \frac{7}{27}$	
H	23	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**: $6, \frac{9}{14}, \frac{9}{245}$	
C	24	Check whether $f(x)=\begin{cases} \frac{3+2x}{18} & ; 2\le x\le 4 \\ 0 & ; \text{otherwise} \end{cases}$ is a Probability density function? If yes, then find $P(3\le X\le 4)$. **Answer**: yes, $\frac{5}{9}$	
H	25	A random variable X has PDF $f(x)=\begin{cases} \frac{3+2x}{18} & ; 2\le x\le 4 \\ 0 & ; \text{otherwise} \end{cases}$. Find the standard deviation of the distribution. **Answer**: 0.5726	
C	26	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**: $\frac{3}{64}, 2.4, 0.8$	
T	27	For the probability function $f(x)=\frac{k}{1+x^2}, -\infty<x<\infty$, find k. **Answer**: $\frac{1}{\pi}$	
T	28	Verify that the following function is pdf or not: $f(x)=\begin{cases} \frac{x}{8} & ; 0\le x<2 \\ \frac{1}{4} & ; 2\le x<4 \\ \frac{6-x}{8} & ; 4\le x<6 \end{cases}$ **Answer**: Yes	

T	29	The life in hours of a certain kind of radio tube has the probability density $f(x) = \begin{cases} \frac{100}{x^2} & ; \text{ for } x \geq 100 \\ 0 & ; \text{ elsewhere} \end{cases}$, find the distribution function and use it to determine the probability that the life of tube is more than 150 hrs. **Answer**: $\mathbf{F(x) = 1 - \frac{100}{x}}$, $\mathbf{P(x > 150) = \frac{2}{3}}$	
H	30	In a certain district, the proportion of highway sections requiring repairs in any given year is a random variable having the probability density $f(x) = \begin{cases} 12x^2(1-x), & \text{for } 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$. Find the distribution function and use it to determine the probability that at least half of the highway's sections will require repairs in any given year. **Answer**: $\mathbf{F(x) = 4x^3 - 3x^4}$, $\mathbf{P\left(x \geq \frac{1}{2}\right) = \frac{11}{16}}$	

❖ 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 outcome. 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 variable. Further,

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

❖ 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

$$P_X(x) = \sum_y P(X = x, Y = y) \quad \& \quad P_Y(y) = \sum_x P(X = x, Y = y).$$

- ✓ **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 Y 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, then

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

- ✓ It is called the joint probability density function of (X, Y), provided $f(x,y)\ge 0$, for all $(x,y)\in D$; Where D is range of space and

$$\iint_D f(x,y)\,dx\,dy=1.\text{ i.e. }P(a\le X\le b, c\le Y\le d)=\int_c^d\int_a^b f(x,y)\,dx\,dy=1.$$

❖ THE MARGINAL PROBABILITY FUNCTION (CONTINUOUS CASE)

- ✓ The marginal probability function is defined as

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

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

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

  - ➢ 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

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

❖ CONDITIONAL DENSITY

- ✓ Conditional density of X:

$f(\,x \mid y) = \dfrac{f(x,y)}{f_X(x)}$, where $f_X(x)$ is marginal probability density function of X and $f(x,y)$ is joint probability density function.

- Conditional density of Y:

  $f(\,y \mid x) = \dfrac{f(x,y)}{f_Y(y)}$, where $f_Y(y)$ is marginal probability density function of Y and $f(x,y)$ is joint probability density function.

INDEPENDENT RANDOM VARIABLES:

- Two random variables X and Y are 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

$$P_X(\,X\ =\ 0) = 0.1 + 0.04 + 0.02 = 0.16.$$

- The marginal Probability Mass Function of Y=0 is

$$P_Y(\,Y\ =\ 0) = 0.1 + 0.08 + 0.06 = 0.24.$$

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

$$\therefore P(X = 0, Y = 0) \neq P_X(0)\,.\,P_Y(0).$$

$\therefore$ 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:

$$E(X) = \iint_R x\, f(x,y)dx\,dy \text{ and } E(Y) = \iint_R y\, f(x,y)dy\,dx \text{ (where R is given region)}$$

METHOD – 6: EXAMPLES ON TWO-DIMENSIONAL RANDOM VARIABLE

H	1	X, Y are two random variables with joint mass function $P(x,y) = \frac{1}{27}(2x + y)$ where $x = 0,1,2$ and $y = 0,1,2$. Find the marginal probabilities. **Answer**: $\mathbf{X: \frac{1}{9}, \frac{1}{3}, \frac{5}{9} \;\&\; Y: \frac{2}{9}, \frac{1}{3}, \frac{4}{9}}$	
C	2	The joint probability mass function is given by $\mathbf{p(x,y) = k(2x + 3y)}$, where $\mathbf{x = 0,1,2}$ and $\mathbf{y = 1,2,3}$. Find **(a)** **k**, **(b)** $\mathbf{P(x \le 1, y \ge 2)}$, **(c)** marginal probability, **(d)** expected value. **Answer**: $\mathbf{(a)\ k = \frac{1}{72},\ (b)\ \frac{17}{36},\ (c)\ X: 0.25, 0.3333, 0.4167,}$ $\mathbf{Y: 0.2083,\ 0.3333,\ 0.4583,\ (d)\ X: 1.1667,\ Y: 2.25}$	
H	3	Let $P(X = 0, Y = 1) = \frac{1}{3}, P(X = 1, Y = -1) = \frac{1}{3}, P(X = 1, Y = 1) = \frac{1}{3}$. Is it the joint probability mass function of X and Y? If yes, find the marginal probability function of X and Y. **Answer**: **Yes**, $\mathbf{P_X(0) = \frac{1}{3},\ P_X(1) = \frac{2}{3} \;\&\; P_Y(-1) = \frac{1}{3},\ P_Y(1) = \frac{2}{3}}$	
H	4	A two-dimensional random variable (X, Y) has 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**: $\mathbf{X: \frac{1}{32}, \frac{3}{32}, \frac{9}{32}, \frac{19}{32} \;\&\; Y: \frac{7}{16}, \frac{9}{16}}$, **No**	

H **5**

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.

	$Y = 1$	$Y = 2$	$Y = 3$	$Y = 4$	$Y = 5$	$Y = 6$
$X = 0$	0	0	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{2}{32}$	$\frac{3}{32}$
$X = 1$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
$X = 2$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{64}$	0	$\frac{2}{64}$

Answer: $\mathbf{\frac{1}{16}, \frac{7}{8}, \frac{23}{64}, \frac{9}{16}}$, **No**

H **6**

The following table represents the joint probability distribution of discrete random variable (X, Y). Find $P(x \le 2, Y = 3), P(X + Y < 4)$ & $P(Y \le 2)$.

	$Y = 1$	$Y = 2$	$Y = 3$	$Y = 4$	$Y = 5$	$Y = 6$
$X = 0$	0	0	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{2}{32}$	$\frac{3}{32}$
$X = 1$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
$X = 2$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{64}$	0	$\frac{2}{64}$

Answer: $\mathbf{\frac{11}{64}, \frac{3}{16}, \frac{3}{16}}$

C **7**

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: $\mathbf{P(0,0) = 0.1515}$, $\mathbf{P(1,0) = 0.2273}$, $\mathbf{P(2,0) = 0.0455}$,
$\mathbf{P(0,1) = 0.3030}$, $\mathbf{P(1,1) = 0.1818}$, $\mathbf{P(0,2) = 0.0909}$

H	8	Three balanced coins are tossed. Let X denote the number of heads on the first two coins and Y denote the number of tails on the last two coins. Find the joint distribution of X and Y. **Answer**: $\mathbf{P(0,1)=\frac{1}{8},\ \ P(0,2)=\frac{1}{8},\ \ P(1,0)=\frac{1}{8}, P(1,1)=\frac{1}{4},}$ $\mathbf{P(1,2)=\frac{1}{8},\ \ P(2,1)=\frac{1}{8},\ \ P(2,2)=\frac{1}{8}}$	
C	9	Check weather $f(x,\ y)=\begin{cases}\frac{1}{8}(6-x-y)\ ; & 0\le x<2,\ \ 2\le y<4\\ 0\ ; & \text{otherwise}\end{cases}$ is probability density function or not? **Answer**: **yes**	
H	10	Let $f(x,y)=\begin{cases}Cxy\ ; & 0<x<4,\ \ 1<y<5\\ 0\ ; & \text{otherwise}\end{cases}$ is the joint density function of two random variables X & Y, then find the value of C. **Answer**: $\mathbf{\frac{1}{96}}$	
C	11	For given P. d. f. $f(x,\ y)=\begin{cases}\frac{3}{4}+xy\ ; & 0\le x<1, 0\le y<1\\ 0\ ; & \text{otherwise}\end{cases}$, find (a) joint probability, (b) marginal probability, (c) $P\left(0\le x\le\frac{1}{2}, 0\le y\le\frac{1}{2}\right)$. **Answer**: $\mathbf{(a) 1,\ \ (b)\ X: \frac{3}{4}+\frac{X}{2}; (0\le x<1)\ \&\ Y: \frac{3}{4}+\frac{Y}{2}; (0\le y<1),}$ $\mathbf{(c)\ \frac{13}{64}}$	
H	12	Suppose, two-dimensional continuous random variable (X, Y) has PDF given by $f(x,y)=\begin{cases}6x^2y\ ; & 0<x<1, 0<y<1\\ 0\ ; & \text{elsewhere}\end{cases}$ (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**: $\mathbf{\frac{3}{8},\ \ \frac{1}{10}}$	

C **13** The joint PDF of a two-dimensional random variable (X, Y) is given by

$$f(x,y)=\begin{cases} 2\,; & 0<x<1, 0<y<x \\ 0\,; & \text{elsewhere} \end{cases}$$

Find the marginal density function of X and Y.

Answer: $\mathbf{f_X(x)=2x;\ 0<x<1\ \ \&\ \ f_Y(y)=2(1-y);\ 0<y<1}$

H **14** Check the independence of X and Y for the following PDF.

$$f(x,y)=\begin{cases} \frac{1}{4}(1+xy) & ;\ -1<x<1, -1<y<1 \\ 0 & ;\ \text{elsewhere} \end{cases}$$

Answer: **No**

C **15** The joint probability density of two random variables is given by

$$f(x_1,x_2)=\begin{cases} 6e^{-2x_1-3x_2} & ;\ \text{for } x_1>0, x_2>0 \\ 0 & ;\ \text{elsewhere} \end{cases}.$$

Find the marginal densities of both the random variables and hence show that the two random variables are independent.

Answer: $\mathbf{f_{X_1}(x_1)=2e^{-2x_1};\ x_1>0\ \ \&\ \ f_Y(y)=3e^{-2x_2};\ x_2>0}$

H **16** The joint probability density of two random variables X_1 and X_2 is given by

$$f(x_1,x_2)=\begin{cases} \frac{1}{5}(x_1+2x_2), & \text{for } 0<x_1<1, 0<x_2<2 \\ 0, & \text{elsewhere} \end{cases}.$$

Find the marginal densities of both the random variables and check whether the two random variables are independent.

Answer: $\mathbf{f_{X_1}(x_1)=\frac{2(x_1+2)}{5};0<x_1<1\ \&\ f_Y(y)=\frac{4x_2+1}{10};0<x_2<2,\text{No}}$

C **17** The random variables X and Y have the following joint probability distribution. What is the expected value of X and Y?

	$Y=0$	$Y=1$	$Y=2$
$X=0$	0.2	0.1	0.2
$X=1$	0	0.2	0.1
$X=2$	0.1	0	0.1

Answer: $\mathbf{0.7,\ \ 1.1}$

C	18	Consider the joint probability density function for X and Y to be $f(x, y) = x^2 y^3; 0 < x < 1 \ \& \ 0 < y < x$, find the expected value of X. **Answer**: $\mathbf{\frac{1}{32}}$	
C	19	If two random variables X and Y have the joint density $f(x, y) = \begin{cases} k(x + y^2), & \text{for } 0 < x < 1, 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases}$ Find k and the mean of the conditional density $f_1(x \mid 0.5)$ where $f_1(x)$ is the marginal probability density of X. **Answer**: $\mathbf{k = \frac{6}{5}, \ \ f_1(x \mid 0.5) = \frac{3}{10}(4x + 1)}$	
H	20	If two random variables X and Y have the joint density $f(x, y) = \begin{cases} k(x^2 + y), & \text{for } 0 < x < 1, 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases}$ Find k and the mean of the conditional density $f_1(x \mid 0.5)$ where $f_1(x)$ is the marginal probability density of X. **Answer**: $\mathbf{k = \frac{6}{5}, \ \ f_1(x \mid 0.5) = \frac{2}{3}(2x^2 + 1)}$	

★ ★ ★ ★ ★ ★ ★ ★ ★

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 distributions can be applied.
- ✓ Probability function of discrete random variable is called probability mass function (P.M.F.) and probability function of continuous random variable is called probability density function (P.D.F.).
- ✓ Some special probability distributions:
  - ➢ Binomial distribution (P.M.F.)
  - ➢ Poisson distribution (P.M.F.)
  - ➢ Normal distribution (P.D.F.)
  - ➢ Exponential distribution (P.D.F.)
  - ➢ Gamma distribution (P.D.F.)

❖ 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
  - ➢ The trials are independent.
  - ➢ Each trial results in only two possible outcomes, labeled as success and failure.
  - ➢ 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 $n = 1, 2, 3, \dots$. The probability mass function of X is

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

- ✓ Examples of Binomial Distribution:
  - ➢ Number of defective bolts in a box containing n bolts.
  - ➢ Number of post-graduates in a group of n people.
  - ➢ Number of oil wells yielding natural gas in a group of n wells test drilled.
  - ➢ In the next 20 births at a hospital. Let X=the number of female births.
  - ➢ Flip a coin 10 times. Let X=number of heads obtained.
- ✓ NOTE:
  - ➢ The mean of binomial distribution is defined as $\mu = E(X) = np$.
  - ➢ The variance of binomial distribution is defined as $V(X) = npq$.
  - ➢ The standard deviation of binomial distribution is defined as $\sigma = \sqrt{npq}$.

METHOD – 1: BASIC EXAMPLES ON BINOMIAL DISTRIBUTION

C	1	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**: $\mathbf{0.0567}$	
C	2	20% Of the bulbs produced are defective. Find probability that at most 2 bulbs out of 4 bulbs are defective. **Answer**: $\mathbf{0.9728}$	
H	3	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. **Answer**: $\mathbf{\frac{27}{64}}$	

C	4	The probability that India wins a cricket test match against Australia is given to be $\frac{1}{3}$. If India and Australia play 3 test 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**: $\mathbf{0.2963}$, $\mathbf{0.7037}$	
H	5	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**: $\mathbf{0.5443}$	
H	6	The probability that in a university, a student will be a post-graduate is 0.6. Determine probability that out of 8 students none, two and at least two will be post-graduate. **Answer**: $\mathbf{0.0007}$, $\mathbf{0.0413}$, $\mathbf{0.9915}$	
H	7	If 20% of the bolts produced by a machine are defective, determine the probability that out of 4 bolts chosen at random, (a) 1, (b) 0, (c) less than 2, bolts will be defective. **Answer**: $\mathbf{0.4096}$, $\mathbf{0.4096}$, $\mathbf{0.8192}$	
H	8	Probability of man hitting a target is $\frac{1}{3}$. If he fires 6 times, what is the probability of hitting (a) at most 5 times? (b) at least 5 times? (c) exactly one? **Answer**: $\mathbf{0.9986}$, $\mathbf{0.0179}$, $\mathbf{0.2634}$	
C	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, (b) exactly two patients, (c) at least two patients, are cured? **Answer**: $\mathbf{0.0039}$, $\mathbf{0.2109}$, $\mathbf{0.9492}$	

H	10	Assume that on 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**: $\mathbf{0.9997}$, $\mathbf{0.0051}$	
C	11	Find the probability that in a family of 4 children there will be at least 1 boy. Assume that the probability of a male birth is 0.5. **Answer**: $\mathbf{0.9375}$	
H	12	Out of 2000 families with 4 children each, how many would you expect to have (a) at least 1 boy, (b) 2 boys, (c) 1 or 2 girls, (d) no girls? Assume equal probabilities for boys and girls. **Answer**: $\mathbf{1875}$, $\mathbf{750}$, $\mathbf{1250}$, $\mathbf{125}$	
C	13	Out of 800 families with 4 children each, how many would you expect to have (a) 2 boys and 2 girls? (b) at least 1 boy? (c) at most 2 girls? (d) no girls? Assume equal probabilities for boys and girls. **Answer**: $\mathbf{300}$, $\mathbf{750}$, $\mathbf{550}$, $\mathbf{50}$	W-19 (7)
T	14	A multiple-choice test consists of 8 questions with 3 answers 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	15	Ten coins are thrown simultaneously. Find the probability of getting at least seven heads. **Answer**: $\mathbf{0.1719}$	
H	16	A dice is thrown 6 times getting an odd number of success. Find probability of (a) five successes, (b) at least five successes, (c) at most five successes. **Answer**: $\mathbf{\frac{3}{32}, \frac{7}{64}, \frac{63}{64}}$	

C	17	Find the probability that in five tosses of a fair die, 3 will appear (a) twice, (b) at most once, (c) at least two times. **Answer**: $\frac{625}{3888}, \frac{3125}{3888}, \frac{763}{3888}$	
C	18	Find probability of getting a sum of 7 at least once in 3 tosses of a pair of dice. **Answer**: $\frac{91}{216}$	
H	19	Find the binomial distribution for $n = 4$ and $p = 0.3$. **Answer**: $P(X = x) = \binom{4}{x}(0.3)^x(0.7)^{4-x}\,; x = 0, 1, 2, 3, 4.$	
C	20	Obtain the binomial distribution for which mean is 10 and variance is 5. **Answer**: $P(X = x) = \binom{20}{x}(0.5)^x(0.5)^{20-x}\,; x = 0, 1, 2, \dots, 20.$	
C	21	For the binomial distribution with $n = 20, p = 0.35$. Find mean, variance and standard deviation. **Answer**: $7,\ \ 4.55,\ \ 2.1331$	
H	22	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**: $40,\ \ 6$	
H	23	A university warehouse has received a shipment of 25 printers, of which 10 are laser printers and 15 are inkjet models. If 6 of these 25 are selected at random to be checked by a particular technician, what is the probability that exactly 3 of those selected are laser printers (so that the other 3 are inkjets)? **Answer**: 0.2765	W-19 (4)
H	24	Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant. Find the probability that in the next 18 samples, at least 4 samples contain the pollutant. **Answer**: 0.0982	W-19 (3)

H	25	The probability that one of the ten telephone lines is busy at an instant is 0.2. (a) What is the probability that 5 of the lines are busy? (b) What is the probability that all the lines are busy? **Answer**: $\mathbf{0.0264}$, $\mathbf{0.0000001}$	
T	26	A safety engineers feels that 30% of all industrial accidents in her plant are caused by failure of employees to follow instructions. If this figure is correct, find approximately, the probability that among 84 industrialized accidents in this plant anywhere from 20 to 30 (inclusive) will be due to failure of employees to follow instructions. **Answer**: $\mathbf{0.8102}$	

❖ POISSON DISTRIBUTION

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

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

- ✓ Examples of Poisson Distribution:
  - ➢ Number of telephone calls per minute at a switchboard.
  - ➢ Number of cars passing a certain point in one minute.
  - ➢ Number of printing mistakes per page in a large text.
  - ➢ 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.
  - ➢ $\lambda = np, \lambda \in (0, \infty)$.

- The mean and variance of the Poisson distribution with parameter λ are defined as follows.

$$\text{mean } \mu = E(X) = \lambda = np \quad \& \quad \text{variance } V(X) = \sigma^2 = \lambda$$

METHOD – 2: EXAMPLES ON POISSON DISTRIBUTION

C	1	In a company, there are 250 workers. The probability of a worker remains absent on any one day is 0.02. Find the probability that on a day, seven workers are absent. **Answer**: $\mathbf{P(X = 7) = 0.1044}$	
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? **Answer**: $\mathbf{0.2642}$	
H	3	Suppose a book of 585 pages contains 43 typographical errors. If these errors are randomly distributed throughout the book, what is the probability that 10 pages, selected at random, will be free from error? **Answer**: $\mathbf{0.4795}$	
H	4	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**: $\mathbf{P(X \leq 3) = 0.8571}$	
C	5	Average number of accidents on any day on a national highway is 1.8. Determine the probability that the number of accidents are (a) at least 1, (b) at most 1. **Answer**: $\mathbf{0.8347}$, $\mathbf{0.4628}$	
H	6	The probability that a person catch corona virus is 0.001. Find the probability that out of 3000 persons (a) exactly 3, (b) more than 2 persons will catch the virus. **Answer**: $\mathbf{0.2240}$, $\mathbf{0.5768}$	

H	7	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 defectives. **Answer**: $\mathbf{P(X=0)=0.3679, P(X=1)=0.3679, P(X\geq 3)=0.0803}$	
C	8	Potholes on a highway can be serious problems. The past experience suggests that there are, on an average, 2 potholes per mile after a certain amount of usage. It is assumed that Poisson process applies to 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**: $\mathbf{P(X\leq 4)=0.0293}$	
H	9	A car hire firm has two cars, which are 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**: $\mathbf{P(X=0)=0.2231,\ \ 1-P(X\leq 2)=0.1912}$	
C	10	In sampling a large number of parts manufactured by a machine, the mean number 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**: $\mathbf{271}$	
C	11	In a certain factory turning out razor blades, there is a small chance of 0.002 for any blade to be defective. The blades are supplied in packet of 10. Calculate the approximate number of packets containing no defective, one defective, two defective blades in a consignment of 10000 packets. **Answer**: $\mathbf{9802,\ \ 196,\ \ 2}$	
H	12	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 20 bolts. Use Poisson distribution to find approximate number of packets containing (a) no defective bolt, (b) containing two defective bolts, in the consignment of 10000 packets. **Answer**: $\mathbf{9608,\ \ 8}$	

C	13	For Poisson variant X, if $P(X = 3) = P(X = 4)$, then find $P(X = 0)$. **Answer**: $\mathbf{P(X = 0) = e^{-4}}$	
H	14	For Poisson variant X, if $P(X = 1) = P(X = 2)$. Find mean and standard deviation of this distribution. Also, find $P(X = 3)$. **Answer**: $\mathbf{2}$, $\mathbf{\sqrt{2}}$, $\mathbf{0.1804}$	
H	15	Assume that the probability that a wafer contains a large particle of contamination is 0.01 and that the wafers are independent; that is, the probability that a wafer contains a large particle is not dependent on the characteristics of any of the other wafers. If 15 wafers are analyzed, what is the probability that no large particles are found? **Answer**: $\mathbf{0.8607}$	W-19 (3)
H	16	If a publisher of nontechnical books takes great pains to ensure that its books are free of typographical errors, so that the probability of any given page containing at least one such error is .005 and errors are independent from page to page. What is the probability that one of its 400-page novels will contain (a) exactly one page with errors? (b) at most three pages with errors? **Answer**: **(a)** $\mathbf{0.2707}$, **(b)** $\mathbf{0.8571}$	W-19 (7)
H	17	If the probability that an individual suffers a bad reaction from a certain injection is 0.001. Find the probability that out of 2000 individuals, (i) more than 2 individuals; (ii) exactly 3 individuals will suffer a bad reaction. **Answer**: $\mathbf{0.3233}$, $\mathbf{0.1804}$	
T	18	The number of flaws in a fiber optic cable follows a Poisson process with an average of 0.6 per 100 feet. (i) Find the probability of exactly 2 flaws in a 200 feet cable. (ii) Find the probability of exactly 1 flaw in the first 100 feet and exactly 1 flaw in the second 100 feet. **Answer**: $\mathbf{0.2169}$, $\mathbf{0.3293}$	

H	19	The number of monthly breakdowns of a computer is a random variable having Poisson distribution with mean 1.8. Find the probability that the computer will function for a month (a) without a breakdown (b) with at least one breakdown. **Answer**: $\mathbf{0.1653,\ \ 0.8347}$	
C	20	The number of page requests that arrive at a Web server is a Poisson random variable. Its probability distribution is as follows: Number of x requests/sec.: 0, 1, 2, 3, 4, 5, 6 Probability f(x): 0.368, 0.368, 0.184, 0.061, 0.015, 0.003, 0.001 Find the mean and variance of this probability distribution. **Answer**: $\mathbf{1,\ \ 1.0004 \cong 1}$	

Number of x requests/sec.	0	1	2	3	4	5	6
Probability f(x)	0.368	0.368	0.184	0.061	0.015	0.003	0.001

❖ EXPONENTIAL DISTRIBUTION

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

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

- ✓ 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.
- ✓ In exponential distribution we can find the probability as given below.
  - ➢ $P(X \leq x) = 1 - e^{-\theta x}$ & $P(X \geq x) = e^{-\theta x}$
  - ➢ $P(a \leq X \leq b) = e^{-a\theta} - e^{-b\theta}$

METHOD – 3: EXAMPLES ON EXPONENTIAL DISTRIBUTION

T	1	Define exponential distribution. Obtain its mean and variance.	

C	2	The lifetime T of an alkaline battery is exponentially distributed with $\theta = 0.05$ per hour. What are mean and standard deviation of batteries lifetime? **Answer**: **20**, **20**	
C	3	The lifetime T of an alkaline battery is exponentially distributed with $\theta = 0.05$ per hour. (a) What are the probabilities for battery to last between 10 and 15 hours? (b) What are the probabilities for the battery to last more than 20 hours? **Answer**: **0.1342**, **0.3679**	
C	4	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-day period. **Answer**: **0.5862**	
C	5	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 of having 3 arrivals in 5 min. interval? **Answer**: **0.0356**, **0.2202**	
T	6	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.? **Answer**: **0.0821**, **0.1481**	
H	7	The time intervals between successive barges passing a certain point on a busy waterway have an exponential distribution with mean 8 minutes. Find the probability that the time interval between two successive barges is less than 5 minutes. **Answer**: **0.4647**	

H	8	Accidents occur with Poisson distribution at an average of 4 per week. (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**: $\mathbf{0.3895}$, $\mathbf{0.0003}$	
C	9	A random variable has an exponential distribution with probability density function given by $f(x) = 3e^{-3x}\,; x > 0$ & $f(x) = 0\,; x \le 0$. What is the probability that X is not less than 4? **Answer**: $\mathbf{e^{-12}}$	
T	10	The income tax of a man is exponentially distributed with $f(x) = \frac{1}{3}e^{-\left(\frac{x}{3}\right)}\,; x > 0$. What is the probability that his income will exceed Rs. 17000? Assume that the income tax is levied at the rate of 15% on the income above Rs. 15000. **Answer**: $\mathbf{e^{-100}}$	

❖ GAMMA DISTRIBUTION

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

$$f(x) = \begin{cases} \dfrac{\theta^r\, x^{r-1}\, e^{-\theta x}}{\Gamma(r)} & ;\ x \ge 0 \\ 0 & ;\ \text{otherwise} \end{cases}$$

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

METHOD – 4: EXAMPLES ON GAMMA DISTRIBUTION

H	1	Define Gamma distribution. Obtain its mean and variance.	
C	2	Given a gamma random variable X with $r = 3$ and $\theta = 2$. Find E(X), V(X) and $P(X \le 1.5)$. **Answer**: $\mathbf{1.5}$, $\mathbf{0.75}$, $\mathbf{0.5768}$	

H	3	The time spent on a computer is a gamma distribution with mean 20 and variance 80. (a) What are the values of r & θ? (b) What is $P(X < 24)$? (c) What is $P(20 < X < 40)$? **Answer**: $\mathbf{r = 5,\ \ \theta = 4,\ \ P(X < 24) = 0.715,\ \ P(20 < X < 40) = 0.411}$	
C	4	The daily consumption of milk in a city, in excess of 20000 liters, is approximately distributed as a gamma variate with $r = 2$ and $\theta = \frac{1}{10000}$. The city has daily stock of 30000 liters. What is the probability that the stock is insufficient on a particular day? **Answer**: $\mathbf{0.736}$	
T	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 θ 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**: $\mathbf{r = 4,\ \ \theta = \frac{1}{2},\ \ 0.3528}$	
H	6	Suppose you are fishing and you expect to get a fish once every $\frac{1}{2}$ hour. Compute the probability that you will have to wait between 2 to 4 hours before you catch 4 fish. **Answer**: $\mathbf{0.1239}$	
C	7	The daily consumption of electric power in a certain city is a random variable X having probability density function $f(x) = \begin{cases} \frac{1}{9}\, x\, e^{-\frac{x}{3}} ; & x > 0 \\ \quad 0 \quad ; & x \le 0 \end{cases}$. 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 per hour. **Answer**: $\mathbf{0.0916}$	

❖ NORMAL DISTRIBUTION

- ✓ A continuous random variable X is said to follows a normal distribution if 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.
- ✓ μ (mean) & σ^2 (variance) are called parameters of the distribution.
- ✓ If X is a normal random variable with mean μ and standard deviation σ, and if we find the random variable $Z = \frac{X-\mu}{\sigma}$ 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

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

- ✓ 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,
  - ➢ $P(-\infty \le z \le \infty) = 1$ (Total area).
  - ➢ $P(-\infty \le z \le 0) = P\,(0 \le z \le \infty) = 0.5$.
  - ➢ $P(-z_1 \le z \le 0) = P(0 \le z \le z_1) \ ; \ z_1 > 0$.

METHOD – 5: EXAMPLES OF NORMAL DISTRIBUTION

C	1	For a random variable having the normal distribution with $\mu = 18.2$ and $\sigma = 1.25$, find the probabilities that it will take on a value (a) less than 16.5, (b) between 16.5 and 18.8.[$P(z = 1.36) = 0.4131, P(z = 0.48) = 0.1843$] **Answer**: $\mathbf{0.0869}$, $\mathbf{0.5974}$	

H	2	The compressive strength of the sample of cement can be modelled by normal distribution with mean $6000\ \mathrm{kg/cm^2}$ and standard deviation $100\ \mathrm{kg/cm^2}$. (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? $[P(z = 2.5) = 0.4798,\ \ P(z = 1) = 0.3413]$ $[P(z = 2) = 0.4773,\ \ P(z = 1.65) = 0.45]$ **Answer**: $\mathbf{0.9798,\ \ 0.136,\ \ 6165}$	
C	3	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. $[P(z = 1.83) = 0.4664,\ \ P(z = 2.33) = 0.4901]$ $[P(z = 0.67) = 0.2486,\ \ P(z = 0.58) = 0.2190]$ **Answer**: $\mathbf{0.9565,\ \ 0.7486,\ \ 0.7190}$	
H	4	A sample of 100 dry battery cell tested & found that average life is 12 hours & standard deviation 3 hours. Assuming data to be normally distributed what % of battery cells are expected to have life (a) more than 15 hrs.? (b) less than 6 hrs.? (c) between 10 & 14 hrs.? $[P(z = 1) = 0.3413,\ \ P(z = 2) = 0.4773,\ \ P(z = 0.67) = 0.2486]$ **Answer**: $\mathbf{15.87\%,\ \ 2.27\%,\ \ 49.72\%}$	
H	5	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? $[P(z = 2) = 0.4773]$ **Answer**: $\mathbf{0.9773}$	

H	6	The customer accounts of certain department store have an average balance of 120 Rs. & Standard deviation of 40 Rs. Assume that account balances are normally distributed. (a) What proportion of the account is over 150 Rs.? (b) What proportion of account is between 100 & 150 Rs.? (c) What proportion of account is between 60 & 90 Rs.? $[P(z = 0.75) = 0.2734, P(z = 0.5) = 0.1915, P(z = 1.5) = 0.4332]$ **Answer: 0.2266, 0.4649, 0.1598**	
C	7	Weights of 500 students of college is normally distributed with $\mu=95$ lbs. & $\sigma=7.5$ lbs. Find how many students will have the weight between 100 and 110 lbs.$[P(z = 2) = 0.4773, P(z = 0.67) = 0.2486]$ **Answer: 114**	
H	8	Distribution of height of 1000 soldiers is normal with mean 165 cm & standard deviation 15 cm. How many soldiers are of height (a) less than 138 cm? (b) more than 198 cm? (c) between 138 & 198 cm? $[P(z = 1.8) = 0.4641, \ P(z = 2.2) = 0.4861]$ **Answer: 36, 14, 950**	W-19 (7)
T	9	Assuming that the diameters of 1000 brass plugs taken consecutively from a machine form a normal distribution with mean 0.7515 cm and standard deviation 0.002 cm. Find the number of plugs likely to be rejected if the approved diameter is 0.752 ± 0.004 cm. $[P(z = 1.75) = 0.4599, \ P(z = 2.25) = 0.4878]$ **Answer: 52**	
T	10	In a company, amount of light bills follows normal distribution with $\sigma = 60$. 11.31% of customers pay bill less than 260. Find average amount of light bill.$[P(z = 1.21) = 0.3869]$ **Answer: 332.60**	
C	11	In a normal distribution, 31% of items are below 45 & 8% are above 64. Determine the mean and standard deviation of this distribution. $[P(z = 0.22) = 0.19, \ P(z = 1.41) = 0.42]$ **Answer:** $\boldsymbol{\mu = 49.9738, \ \sigma = 9.9476}$	

H	12	In an examination, minimum 40 marks for passing and 75 marks for distinction are required. In this examination 45% students passed and 9% obtained distinction. Find average marks and standard deviation of this distribution of marks.[$P(z = 0.125) = 0.05$ and $P(z = 1.34) = 0.41$] **Answer**: $\mathbf{36.40}$, $\mathbf{28.81}$	W-19 (7)

❖ 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,

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

METHOD – 6: EXAMPLES ON 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	If X is a variate such that $E(X) = 3$, $E(X^2) = 13$, show that $P(-2 < X < 8) \geq \frac{21}{25}$.	

H	3	The number of customers who visit a car dealer's showroom on Sunday morning is a random variable with mean 18 and standard deviation 2.5. What is the bound of probability that on Sunday morning the customers will be 8 to 28? **Answer**: $\mathbf{P(8 < X < 28) \geq \frac{15}{16}}$	
C	4	Variate X takes values $-1, 1, 3, 5$ with associate probability $\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{2}$. Compute $p = P\{\lvert x-3 \rvert \geq 1\}$ directly and find an upper bound to 'p' by Chebyshev's inequality. **Answer**: $\mathbf{0.83,\ \ 5.33}$	
T	5	Two unbiased dice are thrown. If X is the sum of the numbers showing up, prove that $P\{\lvert X-7 \rvert \geq 3\} < \frac{35}{54}$. Compare this with actual probability. **Answer**: $\mathbf{\frac{1}{3}}$	
H	6	A random variable X has mean 10, variance 4 and unknown probability distribution. Find 'c' such that $P\{\lvert X-10 \rvert \geq c\} < 0.04$. **Answer**: **10**	

★ ★ ★ ★ ★ ★ ★ ★ ★ ★

UNIT-3 » BASIC STATISTICS

PART-I CENTRAL TENDENCY AND DISPERSION

❖ INTRODUCTION

- ✓ Statistics is the branch of science where we plan, gather and analyze information about a particular collection of objects 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 characteristic or they differ from each other in the following ways:
  - ➢ They show a tendency to concentrate values, usually somewhere in the center 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.**

❖ 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
  - ➢ Central Tendency
  - ➢ Dispersion

- Skewness
- Kurtosis

❖ 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
  - Distribution of grouped data
    1) Discrete Frequency Distribution
    2) Continuous Frequency Distribution

❖ EXAMPLE FOR DISTRIBUTION

- ✓ Consider the marks obtained by 10 students in a mathematics test as given below:

55, 36, 95, 73, 60, 42, 25, 78, 75, 62

  - The data in this form is called ungrouped data.
  - Let us arrange the marks in ascending order as: 25, 36, 42, 55, 60, 62, 73, 75, 78, 95
  - We can clearly see that the lowest marks are 25 and the highest marks are 95. The difference of the highest and the lowest values in the data is called the range of the data. So, the range in this case is 95 – 25 = 70.
- ✓ Consider the marks obtained (out of 100 marks) by 30 students of Class-XII of a school:

10, 20, 36, 92, 95, 40, 50, 56, 60, 70, 92, 88, 80, 70, 72
70, 36, 40, 36, 40, 92, 40, 50, 50, 56, 60, 70, 60, 60, 88

  - Recall that the number of students who have obtained a certain number of marks is called the frequency of those marks. For instance, 4 students got 70 marks. So the frequency of 70 marks is 4. To make the data more easily understandable, we write it in a table, as given below:

x	10	20	36	40	50	56	60	70	72	80	88	92	95
f	1	1	3	4	3	2	4	4	1	1	2	3	1

  - Above distribution is called the discrete frequency distribution.

✓ 100 plants each were planted in 100 schools during Van Mahotsav. After one month, the number of plants that survived were recorded as:

95	67	28	32	65	65	69	33	98	96
76	42	32	38	42	40	40	69	95	92
75	83	76	83	85	62	37	65	63	42
89	65	73	81	49	52	64	76	83	92
93	68	52	79	81	83	59	82	75	82
86	90	44	62	31	36	38	42	39	83
87	56	58	23	35	76	83	85	30	68
69	83	86	43	45	39	83	75	66	83
92	75	89	66	91	27	88	89	93	42
53	69	90	55	66	49	52	83	34	36

- The Coefficient of variation is lesser is said to be less variable or more consistent. To present such a large amount of data so that a reader can make sense of it easily, we condense it into groups like 20.5-29.5, 29.5-39.5, ..., 89.5-99.5. (Since our data is from 23 to 98)
- These groupings are called 'classes' or 'class-intervals', and their size is called the class-size or class width, which is 10 in this case. In each of these classes, the least number is called the lower-class limit and the greatest number is called the upper-class limit, e.g., in 20-29, 20 is the 'lower class limit' and 29 is the 'upper class limit'.
- Also, recall that using tally marks, the data above can be condensed in tabular form as follows:

Class	22.5-29.5	29.5-39.5	39.5-49.5	49.5-59.5	59.5-69.5	69.5-79.5	79.5-89.5	89.5-99.5
f	3	14	12	8	18	10	23	12

- Above distribution is called the continuous frequency distribution.

❖ SOME DEFINITION

✓ Exclusive Class: If classes of frequency distributions are $0 - 2, 2 - 4, 4 - 6, \ldots$ such classes are called Exclusive Classes.

✓ Inclusive Class: If classes of frequency distributions are $0 - 2, 3 - 5, 6 - 8, \ldots$ such classes are called Inclusive Classes.

- ✓ Mid-Point of class: It is defined as

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

❖ 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 ($\bar{x}$)
  - ➢ Median (M)
  - ➢ Mode(Z)

❖ MEAN

- ✓ 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 of Ungrouped Data

$$\bar{x} = \frac{\sum x_i}{n}$$

- ✓ Mean of Discrete Grouped Data

$$\bar{x} = \frac{\sum f_i x_i}{n}$$

  - ➢ Mean by assumed mean method: $\bar{x} = A + \frac{\sum f_i d_i}{n}$; Where $d_i = x_i - A$
  - ➢ Here, A can be any value of x_i.
- ✓ Mean of Continuous Grouped Data

$$\bar{x} = \frac{\sum f_i x_i}{n} \text{ ; Where } x_i = \text{Mid value of the respective class}$$

  - ➢ Mean by assumed mean method: $\bar{x} = A + \frac{\sum f_i d_i}{n}$; Where $d_i = x_i - A$
  - ➢ Mean by step deviation method: $\bar{x} = A + \frac{\sum f_i u_i}{n} \cdot C$; Where $u_i = \frac{(x_i - A)}{C}$
  - ➢ Here, A can be any value of x_i.
  - ➢ C is the class length.

METHOD-1: EXAMPLES ON MEAN

H **1** Find the mean for following data:

(a) 10.2, 9.5, 8.3, 9.7, 9.5, 11.1, 7.8, 8.8, 9.5, 10. (b) −1.5, 0, 1, 0.8. (c) 2, 8, 4, 6, 10, 12, 4, 8, 14, 16. (d) 10, 9, 21, 16, 14, 18, 20, 18, 14, 18, 23, 16, 18, 4.

Answer: $\mathbf{9.44, 0.075, 8.4, 15.6429}$

C **2** Find the mean for following data:

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

Answer: $\mathbf{32.23}$

H **3** Find the mean for following data:

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

Answer: $\mathbf{34.5}$

H **4** Find the mean for following data:

x	10	20	36	40	50	56	60	70	72	80	88	92	95
f	1	1	3	4	3	2	4	4	1	1	2	3	1

Answer: $\mathbf{59.3}$

C **5** The following data represents the no. of foreign visitors in a multinational company in every 10 days during last 2 months. Use the data find to the mean.

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

Answer: $\mathbf{28}$

H **6** Find the mean if Survey regarding the weights (kg) of 45 students of class X of a school was conducted and the following data was obtained:

Weight (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

Answer: $\mathbf{38.83}$

C **7** Find the mean using direct method, assumed mean method and step deviation method:

Marks	0-10	10-20	20-30	30-40	40-50
No. of students	5	10	40	20	25

Answer: **30**

T **8** Find the missing frequency from the following data if mean is 19.92.

x	4-8	8-12	12-16	16-20	20-24	24-28	28-32	32-36	36-40
f	11	13	16	14	?	9	17	6	4

Answer: **10**

C **9** A co-operative bank has two branches employing 50 and 70 workers respectively. The average salaries paid by two respective branches are 360 and 390 rupees per month. Calculate the mean of the salaries of all the employees.

Answer: **377.5**

T **10** A car runs at speed of 60 k/h over 50 km; the next 30 km at speed of 40 k/h; next 20 km at speed of 30 k/h; final 50 km at speed of 25 k/h. What is the average speed?

Answer: **35.29**

❖ MEDIAN

- ✓ 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
  - ➢ If n is odd number, then

$$M = \left(\frac{n+1}{2}\right)^{th} \text{ observation}$$

  - ➢ If n is even number, then

$$M = \frac{\left(\frac{n}{2}\right)^{th} \text{observation} + \left(\frac{n}{2}+1\right)^{th} \text{ observation}}{2}$$

- ✓ 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

$$M = L + \left(\frac{\frac{n}{2} - F}{f}\right) \times C$$

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

METHOD-2: EXAMPLES ON MEDIAN

H 1 Find the median of following data:

(a) 20, 25, 30, 15, 17, 35, 26, 18, 40, 45, 50.

(b) 110, 115, 108, 112, 120, 116, 140, 135, 128, 132.

(c) 6, 20, 43, 50, 19, 53, 0, 37, 78, 1, 15.

(d) 10, 34, 27, 24, 12, 27, 20, 18, 15, 30.

Answer: 26, 118, 20, 22

C 2 Calculate the median for the following data:

Marks	20	9	25	50	40	80
No. of students	6	4	16	7	8	2

Answer: 25

H **3** Obtain the median size of shoes sold from the following data:

Size	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10	10.5	11	11.5
Pair	30	40	50	150	300	600	950	820	750	440	250	150	40	39

Answer: 8.5

H **4** Obtain median for the following frequency distribution:

x	1	2	3	4	5	6	7	8	9
f	8	10	11	16	20	25	15	9	6

Answer: 5

C **5** The following table gives marks obtained by 50 students in statistics. Find the median.

Marks	0-10	10-20	20-30	30-40	40-50
No. of students	16	12	18	3	1

Answer: 17.5

H **6** An insurance company obtained data for accident claims from a region. Find median.

Amount of claim (thousand)	1-3	3-5	5-7	7-9	9-11	11-13	13-15	15-17
Frequency	6	53	85	56	21	16	4	4

Answer: 6.49

C **7** The given observations have been arranged in ascending order. If the median of the data is 63, find the value of x: 29, 32, 48, 50, x, x + 2, 72, 78, 84, 95.

Answer: 62

T **8** The median of 60 observations (following data) is 28.5. Find x and y.

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	5	x	20	15	y	5

Answer: 8, 7

C **9** The following table gives the marks obtained by 50 students in mathematics. Find the median.

Marks	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
No. of students	4	6	10	5	7	3	9	6

Answer: 29.5

❖ MODE

- ✓ 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 Discrete Frequency Distribution
  - ➢ The value of variable corresponding to highest frequency.
- ✓ Mode of Continuous Frequency Distribution

$$Z = L + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times C$$

  - ➢ The modal class is the class with highest frequency.
  - ➢ 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

METHOD-3: EXAMPLES ON MODE

H **1** Find the mode of following data:

(a) 2, 4, 2, 5, 7, 2, 8, 9.

(b) 2, 8, 4, 6, 10, 12, 4, 8, 14, 16.

Answer: **2**, **4** & **8**

H **2** Find the mode of following data:

x	1	2	3	4	5	6	7	8	9
f	8	10	11	16	20	25	15	9	6

Answer: **6**

C **3** Find the mode of following data:

x	11	22	33	44
f	15	20	19	10

Answer: **22**

T **4** Find the mode from the following frequency distribution:

x	8	9	10	11	12	13	14	15
f	5	6	8	7	9	8	9	6

Answer: **12**, **14**

C **5** Find the mode of following data:

class	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
f	3	5	7	10	12	15	12	6	2	8

Answer: **55**

H **6** Find the mode of following data:

x	200-220	220-240	240-260	260-280	280-300	300-320	320-340
f	7	15	21	19	6	4	2

Answer: **255**

<table>
<tr><td>H</td><td>7</td><td>Find the mode of following data:<table><tr><td>x</td><td>400-500</td><td>500-600</td><td>600-700</td><td>700-800</td><td>800-900</td></tr><tr><td>f</td><td>8</td><td>16</td><td>20</td><td>17</td><td>3</td></tr></table>**Answer: 657.14**</td><td></td></tr>
<tr><td>C</td><td>8</td><td>In an asymmetrical distribution mean is 16 & median is 20. Calculate the mode.
Answer: 28 [Hint: Use $Z = 3M - 2\bar{x}$]</td><td></td></tr>
</table>

❖ 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** 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 36 - 15 = 21.
- ✓ **Standard Deviation (σ)** is a measure that is used to quantify the amount of variation or dispersion of a set of data values.
- ✓ **Variance ($V = \sigma^2$)** is expectation of the squared deviation. It informally measures how far a set of (random) numbers are spread out from their mean.
- ✓ **Coefficient of variation** is defined and denoted by

$$C.V. = \frac{\sigma}{\bar{x}} \times 100$$

  - ➢ The Coefficient of variation is lesser is said to be less variable or more consistent.

- ✓ **Sample Standard Deviation (S)** is root mean square of the difference between observation and sample mean. It is defined by

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}, \text{ where } \bar{x} \text{ is a sample mean}$$

- ✓ **Sample Variance (S^2)** is the average of squared difference from the mean. It is defined by

$$(S^2) = \frac{\sum (x_i - \bar{x})^2}{n - 1}, \text{ where } \bar{x} \text{ is a sample mean}$$

- ✓ Table of different formulas of standard deviation

Method	Ungrouped Data	Grouped Data
Direct Method	$\sigma=\sqrt{\frac{\sum x_i^2}{n}-\left(\frac{\sum x_i}{n}\right)^2}$	$\sigma=\sqrt{\frac{\sum f_i x_i^2}{n}-\left(\frac{\sum f_i x_i}{n}\right)^2}$
Actual Mean Method	$\sigma=\sqrt{\frac{\sum (x_i-\bar{x})^2}{n}}$	$\sigma=\sqrt{\frac{\sum f_i (x_i-\bar{x})^2}{n}}$
Assumed Mean Method	$\sigma=\sqrt{\frac{\sum d_i^2}{n}-\left(\frac{\sum d_i}{n}\right)^2}$	$\sigma=\sqrt{\frac{\sum f_i d_i^2}{n}-\left(\frac{\sum f_i d_i}{n}\right)^2}$
Step Deviation Method	$\sigma=\sqrt{\frac{\sum u_i^2}{n}-\left(\frac{\sum u_i}{n}\right)^2}\times c$	$\sigma=\sqrt{\frac{\sum f_i u_i^2}{n}-\left(\frac{\sum f_i u_i}{\sum nf_i}\right)^2}\times c$

- ➢ For assumed mean method: $d_i = x_i - A$.
- ➢ For step deviation method: $u_i = \frac{x_i-A}{c}$.

- ✓ Table of different formulas of mean deviation

Method	Ungrouped Data	Grouped Data
M.D. about Mean	$\mathrm{M.D.}=\frac{\sum\lvert x_i-\bar{x}\rvert}{n}$	$\mathrm{M.D.}=\frac{\sum f_i\lvert x_i-\bar{x}\rvert}{n}$
M.D. about Median	$\mathrm{M.D.}=\frac{\sum\lvert x_i-M\rvert}{n}$	$\mathrm{M.D.}=\frac{\sum f_i\lvert x_i-M\rvert}{n}$
M.D. about Mode	$\mathrm{M.D.}=\frac{\sum\lvert x_i-Z\rvert}{n}$	$\mathrm{M.D.}=\frac{\sum f_i\lvert x_i-Z\rvert}{n}$

METHOD-4: EXAMPLES ON DISPERSION

C	1	Find the standard deviation for the following data: 6, 7, 10, 12, 13, 4, 8, 12. **Answer**: **3.0414**	

<table>
<tr><td>C</td><td>2</td><td>Find the standard deviation and variance for the following distribution:
<table>
<tr><td>x</td><td>0-10</td><td>10-20</td><td>20-30</td><td>30-40</td><td>40-50</td><td>50-60</td><td>60-70</td></tr>
<tr><td>f</td><td>6</td><td>14</td><td>10</td><td>8</td><td>1</td><td>3</td><td>8</td></tr>
</table>
Answer: 19.6214, 384.9993</td><td></td></tr>
<tr><td>T</td><td>3</td><td>Find the standard deviation for the following distribution:
<table>
<tr><td>x</td><td>0-100</td><td>100-200</td><td>200-300</td><td>300-400</td><td>400-500</td><td>500-600</td><td>600-700</td><td>700-800</td></tr>
<tr><td>f</td><td>6</td><td>10</td><td>18</td><td>20</td><td>15</td><td>12</td><td>10</td><td>9</td></tr>
</table>
Answer: 196.21</td><td></td></tr>
<tr><td>H</td><td>4</td><td>Find the standard deviation and variance of the mark distribution of 30 students at mathematics examination in a class as below:
<table>
<tr><td>x</td><td>10-25</td><td>25-40</td><td>40-55</td><td>55-70</td><td>70-85</td><td>85-100</td></tr>
<tr><td>f</td><td>2</td><td>3</td><td>0</td><td>14</td><td>8</td><td>3</td></tr>
</table>
Answer: 19.3391, 374.0008</td><td></td></tr>
<tr><td>H</td><td>5</td><td>The article “A Thin-Film Oxygen Uptake Test for the Evaluation of Automotive Crankcase Lubricants” reported the following data on oxidation-induction time (min) for various commercial oils:
87, 103, 130, 160, 180, 195, 132, 145, 211,
105, 145, 153, 152, 138, 87,99, 93, 119, 129
(i) Calculate the sample variance and standard deviation.
(ii) If the observations were re-expressed in hours, what would be the resulting values of the sample variance and sample standard deviation?
Answer: 1198.1982, 34.6150, 1264.7660, 35.5635</td><td>W-19
(4)</td></tr>
<tr><td>C</td><td>6</td><td>Runs scored by two batsmen A, B in 9 consecutive matches are given below:
<table>
<tr><td>A</td><td>85</td><td>20</td><td>62</td><td>28</td><td>74</td><td>5</td><td>69</td><td>4</td><td>13</td></tr>
<tr><td>B</td><td>72</td><td>4</td><td>15</td><td>30</td><td>59</td><td>15</td><td>49</td><td>27</td><td>26</td></tr>
</table>
Which of the batsman is more consistent?
Answer: B</td><td></td></tr>
</table>

H **7** Two machines A, B are used to fill a mixture of cement concrete in a beam. Find the standard deviation of each machine & comment on the performances of two machines.

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

Answer: $\boldsymbol{\sigma_A = 25.4950}$, $\boldsymbol{\sigma_B = 24.4290}$. **There is less variability in the performance of the machine B.**

H **8** Goals scored by two team A and B in a football season were as shown in the table. Find out which team is more consistent?

Number of goals in a match	0	1	2	3	4
Team A	27	9	8	5	4
Team B	17	9	6	5	3

Answer: **B**

C **9** The arithmetic means of runs scored by three batsmen A, B and C, in the same series of 10 innings, are 50, 48 and 12 respectively. The standard deviations of their runs are 15, 12 and 2 respectively. Who is the most consistent of the three?

Answer: **C**

H **10** The runs scored by two batsmen A and B in 10 matches are given in the following table:

A	14	13	26	53	17	29	79	36	84	49
B	37	22	56	52	14	10	37	48	20	4

Who is more consistent?

Answer: **B**

H **11** Find the mean deviation about the mean for the following data:

12, 3, 18, 17, 4, 9, 17, 19, 20, 15, 8, 17, 2, 3, 16, 11, 3, 1, 0, 5.

Answer: **6.2**

<table>
<tr><td>C</td><td>12</td><td>Find mean deviation about the mean for the following data:
<table><tr><td>x</td><td>2</td><td>5</td><td>6</td><td>8</td><td>10</td><td>12</td></tr><tr><td>f</td><td>2</td><td>8</td><td>10</td><td>7</td><td>8</td><td>5</td></tr></table>Answer: 2.3</td><td></td></tr>
<tr><td>H</td><td>13</td><td>Find mean deviation about the mean for the following data:
<table><tr><td>x</td><td>5</td><td>10</td><td>15</td><td>20</td><td>25</td></tr><tr><td>f</td><td>7</td><td>4</td><td>6</td><td>3</td><td>5</td></tr></table>Answer: 6.3200</td><td></td></tr>
<tr><td>C</td><td>14</td><td>Find out mean deviation about median for the following series:
<table><tr><td>Size</td><td>4</td><td>6</td><td>8</td><td>10</td><td>12</td><td>14</td><td>16</td></tr><tr><td>Freq.</td><td>2</td><td>1</td><td>3</td><td>6</td><td>4</td><td>3</td><td>1</td></tr></table>Answer: 2.4</td><td></td></tr>
<tr><td>H</td><td>15</td><td>Find out mean deviation about median for the following series:
<table><tr><td>Size</td><td>4</td><td>6</td><td>8</td><td>10</td><td>12</td><td>14</td><td>16</td></tr><tr><td>Freq.</td><td>1</td><td>2</td><td>4</td><td>5</td><td>4</td><td>3</td><td>1</td></tr></table>Answer: 2.4</td><td></td></tr>
</table>

PART-II MOMENTS, SKEWNESS AND KURTOSIS

❖ MOMENTS

- ✓ Moment is a familiar mechanical term which refer to the measure of a force respect to its tendency to provide rotation or 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 Distributions
  - ➢ The moments about actual mean

$$\mu_r = \frac{\Sigma\, f_i\,(x-\bar{x})^r}{n}\ ;\ r = 1, 2, 3, 4, \ldots$$

  - ➢ The moments about assumed mean

$$\mu'_r = \frac{\Sigma\, f_i\,(x-a)^r}{n}\ ; r = 1, 2, 3, 4, \ldots$$

  - ➢ The moments about zero

$$v_r = \frac{\Sigma f_i\, x^r}{n}\ ; r = 1, 2, 3, 4, \ldots$$

- ✓ Relation between moments
  - ➢ Moments about actual mean in terms of moments about assumed mean (μ in μ')

$$\mu_1 = \mu'_1 - \mu'_1 = 0$$

$$\mu_2 = \mu'_2 - (\mu'_1)^2$$

$$\mu_3 = \mu'_3 - 3\mu'_2\ .\ \mu'_1 + 2(\mu'_1)^3$$

$$\mu_4 = \mu'_4 - 4\mu'_3\ .\ \mu'_1 + 6\mu'_2\ .(\mu'_1)^2 - 3(\mu'_1)^4$$

  - ➢ Moments about zero using μ & μ'

$$v_1 = a + \mu'_1 = \bar{x}$$

$$v_2 = \mu_2 + (v_1)^2$$

$$v_3 = \mu_3 + 3v_1v_2 - 2{v_1}^3$$

$$v_4 = \mu_4 + 4v_1v_3 - 6{v_1}^2v_2 + 3{v_1}^4$$

- ✓ NOTE: The first moment about zero(v_1) is **MEAN** of data, the second moment about actual mean(μ_2) is **VARIANCE** of data, third moment about actual mean(μ_3) is use to find **SKEWNESS** and forth moments about actual mean(μ_4) is use to find **KURTOSIS**.

❖ SKEWNESS

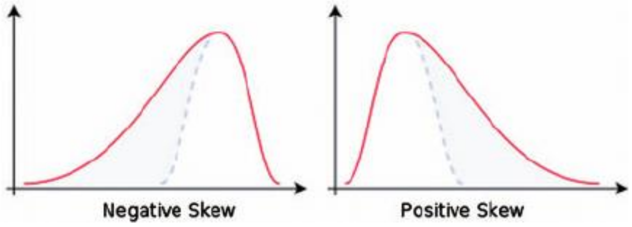
- ✓ It is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean.
  - ➢ Karl Pearson's method

$$\text{Skewness} = \frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

  - ➢ Method of moments

$$\text{Skewness} = \beta_1 = \frac{(\mu_3)^2}{(\mu_2)^3}$$

- ✓ NOTE: The Skewness value can be positive or negative, or even undefined.
  - ➢ **Negative:** The left tail is longer; the mass of the distribution is concentrated on the right.
  - ➢ **Positive:** The right tail is longer; the mass of the distribution is concentrated on the left



  - ➢ If skewness value is zero, then the distribution is called symmetric.

❖ 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

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

- ✓ 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.

METHOD-5: EXAMPLES ON MOMENTS, SKEWNESS AND KURTOSIS

<table>
<tr><td>C</td><td>1</td><td>(a) Find the first four moments about the mean for data 1, 3, 7, 9, 10.
(b) Find the first four central moments for the data 11, 12, 14, 16, 20.
Answer: $[\mathbf{0},\ \mathbf{12},\ \mathbf{-12},\ \mathbf{208.8}],\ [\mathbf{0},\ \mathbf{10},\ \mathbf{19.1520},\ \mathbf{213.5872}]$</td><td></td></tr>
<tr><td>C</td><td>2</td><td>Calculate the first four moments about the mean for following distribution:
<table>
<tr><td>x</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr>
<tr><td>f</td><td>1</td><td>3</td><td>7</td><td>3</td><td>1</td></tr>
</table>
Answer: $[\mathbf{0},\ \mathbf{0.933},\ \mathbf{0},\ \mathbf{2.533}]$</td><td></td></tr>
<tr><td>H</td><td>3</td><td>Calculate the first four moments about the mean for the following data:
<table>
<tr><td>x</td><td>5</td><td>10</td><td>15</td><td>20</td><td>25</td></tr>
<tr><td>f</td><td>6</td><td>10</td><td>14</td><td>6</td><td>4</td></tr>
</table>
Answer: $[\mathbf{0},\ \mathbf{34},\ \mathbf{409.5},\ \mathbf{2702.95}]$</td><td></td></tr>
<tr><td>H</td><td>4</td><td>Calculate the first four moments about the mean of the following data:
<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: $[\mathbf{0},\ \mathbf{2},\ \mathbf{0},\ \mathbf{11}]$</td><td></td></tr>
<tr><td>H</td><td>5</td><td>Calculate the moments about actual mean and zero for following distribution:
<table>
<tr><td>x</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr>
<tr><td>f</td><td>5</td><td>4</td><td>3</td><td>7</td><td>1</td><td>1</td></tr>
</table>
Answer: $[\mathbf{2.90476},\ \mathbf{10.52381},\ \mathbf{43.19048},\ \mathbf{191.66667}]$,
$[\mathbf{0},\ \mathbf{2.08617},\ \mathbf{0.50167},\ \mathbf{9.02988}]$</td><td></td></tr>
</table>

H **6** (a) Calculate the first four moments about the mean for following distribution:

x	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f	8	12	20	30	15	10	5

(b) Calculate the first four moments about the mean for following distribution:

x	60-62	63-65	66-68	69-71	72-74
f	5	18	42	27	8

Answer: $[\mathbf{0}, \mathbf{236.76}, \ \mathbf{264.336}, \ \mathbf{141290}]$,
$[\mathbf{0}, \mathbf{8.5275}, \ \mathbf{-2.6932}, \ \mathbf{199.3759}]$

C **7** Calculate the moments about assumed mean 25, actual mean and zero for following:

x	0-10	10-20	20-30	30-40
f	1	3	4	2

Answer: $[\mathbf{-3}, \ \mathbf{90}, \ \mathbf{-900}, \ \mathbf{21000}]$, $[\mathbf{0}, \ \mathbf{81}, \ \mathbf{-144}, \ \mathbf{14817}]$,
$[\mathbf{22}, \ \mathbf{565}, \ \mathbf{15850}, \ \mathbf{471625}]$

C **8** The first four moments about a = 4 are 1, 4, 10, 45. Find moments about actual mean.

Answer: $\mathbf{0}, \ \mathbf{3}, \ \mathbf{0}, \ \mathbf{26}$

H **9** The first four moment about a = 5 are −4, 22, −117, 560. Find moments about actual mean and origin.

Answer: $[\mathbf{0}, \ \mathbf{6}, \ \mathbf{19}, \ \mathbf{32}]$, $[\mathbf{1}, \ \mathbf{7}, \ \mathbf{38}, \ \mathbf{145}]$

C **10** Karl Pearson's coefficient of skewness of a distribution is 0.32, its standard deviation is 6.5 and mean is 29.6. Find the mode of the distribution.

Answer: $\mathbf{27.52}$

C **11** (a) Compute the coefficient of skewness for the data; 25, 15, 23, 40, 27, 25, 23, 25, 20.

(b) The pH of a solution is measured 7 times by one operator using a same instrument are 7.15, 7.20, 7.18, 7.19, 7.21, 7.16 and 7.18. Find skewness.

Answer: $\mathbf{-0.03}, \ \mathbf{0.0496}$

C **12** (a) Show that the below distribution is symmetric using coefficient of skewness.

x	2	3	4	5	6
f	1	3	7	3	1

(b) Find skewness from the following table.

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

Answer: $\mathbf{0.5788}$

H **13** Find the skewness of the data given below:

x	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f	4	8	3	20	3	4	8

Answer: $\mathbf{0.0445}$

H **14** Find Karl Pearson's coefficient of skewness for the following data:

x	0-10	10-20	20-30	30-40	40-50
f	13	20	30	25	12

Answer: $\mathbf{-0.1135}$

H **15** Find skewness by the method of moments for 38.2, 40.9, 39.5, 44, 39.6, 40.5, 39.5.

Answer: $\mathbf{0.4261}$

H **16** Find skewness by the method of moments for the following frequency distribution:

x	2	3	4	5	6
f	1	3	7	3	1

Answer: $\mathbf{0}$

C **17** Find skewness by the method of moments for the following frequency distribution:

x	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f	4	8	3	20	3	4	8

Answer: $\mathbf{0.0445}$

H **18** Find the coefficient of skewness based on the method of moments for the following data:

Class	0-10	10-20	20-30	30-40	40-50
f	13	20	30	25	12

Answer: -0.1135

H **19** Show that the kurtosis of the data given data is 2.102.

x	5	15	25	35
f	1	4	3	2

C **20** Find out the kurtosis of the data given below:

x	0-10	10-20	20-30	30-40	40-50
f	10	20	40	20	10

Answer: 2.5

PART-III CORRELATION AND REGRESSION

❖ 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 lying 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
  - ➢ If both the variables vary in the same direction, the correlation is said to be positive.
  - ➢ If both the variables vary in the opposite direction, correlation is said to be negative.
- ✓ Simple, partial and multiple correlations
  - ➢ When only two variables are studied, the relationship is described as simple correlation.
  - ➢ When more than two variables are studied, the relationship is multiple correlation.
  - ➢ When more than two variables are studied excluding some other variables, the relationship is termed as partial correlation.
- ✓ Linear and nonlinear correlations
  - ➢ If the ratio of change between two variables is constant, the correlation is said to be linear.

- If the ratio of change between two variables is not constant, the correlation is nonlinear.

MATHEMATICAL METHODS OF STUDYING CORRELATION

- ✓ Karl Pearson's coefficient of correlation (r)

$$r = \frac{\Sigma\,(x - \bar{x})\,(y - \bar{y})}{n\,\sigma_x\,\sigma_y} \quad \textbf{OR} \quad r = \frac{n\,\Sigma\,x\,y - (\Sigma\,x)\,(\Sigma\,y)}{\sqrt{n\,\Sigma\,x^2 - (\Sigma\,x)^2}\,\sqrt{n\,\Sigma\,y^2 - (\Sigma\,y)^2}}$$

- ✓ Spearman's rank coefficient of correlation (ρ)
  - 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(ρ) is given by

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

  - Where d is difference between the ranks R_1 & R_2 given by two judges, n = number of pairs.
  - If there is a tie between individuals' ranks, the rank is divided among equal individuals.
  - For example, if two items have fourth rank, the 4th and 5th rank is divided between them equally and is given as $\frac{4+5}{2} = 4.5^{th}$ rank to each of them.
  - If three items have the same 4th rank, each of them is given $\frac{4+5+6}{3} = 5^{th}$ rank.
  - So, if m is number of items having equal ranks, then the factor $\left(\frac{1}{12}\right)(m^3 - m)$ is added to $\sum d^2$. If there are more than one cases of these types, this factor is added corresponding to each case.

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

  - If there is a tie between the ranks, then it is known as Tied rank.

METHOD-6: EXAMPLES ON CORRELATION

<table>
<tr><td>C</td><td>1</td><td>Calculate the co-efficient of correlation between the given series:
<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>Compute the coefficient of correlation between X and Y using the following data:
<table>
<tr><td>x</td><td>2</td><td>4</td><td>5</td><td>6</td><td>8</td><td>11</td></tr>
<tr><td>y</td><td>18</td><td>12</td><td>10</td><td>8</td><td>7</td><td>5</td></tr>
</table>
Answer: $\mathbf{r = -0.9203}$</td><td>W-19
(3)</td></tr>
<tr><td>H</td><td>3</td><td>Compute Karl Pearson’s coefficient of correlation between X and Y for the following data:
<table>
<tr><td>X</td><td>100</td><td>98</td><td>78</td><td>85</td><td>110</td><td>93</td><td>80</td></tr>
<tr><td>Y</td><td>85</td><td>90</td><td>70</td><td>72</td><td>95</td><td>81</td><td>74</td></tr>
</table>
Answer: $\mathbf{0.9603}$</td><td></td></tr>
<tr><td>H</td><td>4</td><td>Calculate the coefficient of correlation for the following series:
<table>
<tr><td>x</td><td>65</td><td>66</td><td>67</td><td>67</td><td>68</td><td>69</td><td>70</td><td>72</td></tr>
<tr><td>y</td><td>67</td><td>68</td><td>65</td><td>68</td><td>72</td><td>72</td><td>69</td><td>71</td></tr>
</table>
Answer: $\mathbf{r = 0.6030}$</td><td></td></tr>
<tr><td>H</td><td>5</td><td>Calculate the coefficient of correlation for the following series:
<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>H</td><td>6</td><td>Calculate the coefficient of correlation between the age of husband and wife for below:
<table>
<tr><td>Age of husband</td><td>35</td><td>34</td><td>40</td><td>43</td><td>56</td><td>20</td><td>38</td></tr>
<tr><td>Age of wife</td><td>32</td><td>30</td><td>31</td><td>32</td><td>53</td><td>20</td><td>33</td></tr>
</table>
Answer: $\mathbf{r = 0.9371}$</td><td></td></tr>
</table>

T **7** Calculate Karl-Pearson's correlation coefficient between age and playing habits:

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: $\mathbf{r = -0.9738}$

H **8** Find the correlation coefficient between the serum diastolic B.P. and serum cholesterol levels of 10 randomly selected data of 10 persons.

Person	1	2	3	4	5	6	7	8	9	10
Cholesterol	307	259	341	317	274	416	267	320	274	336
B.P.	80	75	90	74	75	110	70	85	88	78

Answer: $\mathbf{0.8088}$

C **9** Determine the coefficient of correlation if $n = 10, \bar{x} = 5.5, \bar{y} = 4, \sum x^2 = 385, \sum y^2 = 192, \sum (x+y)^2 = 947$.

Answer: $\mathbf{r = -0.6812}$

H **10** Determine the coefficient of correlation if $n = 8, \bar{x} = 0.5, \bar{y} = 0.5, \sum x^2 = 44, \sum y^2 = 44, \sum xy = -40$.

Answer: $\mathbf{r = -1}$

C **11** Find r_{xy} from given data if $n = 10, \sum (x-\bar{x})(y-\bar{y}) = 66, \sigma_x = 5.4, \sigma_y = 6.2$.

Answer: $\mathbf{r = 0.1971}$

H **12** 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: $\mathbf{r = 0.7857}$

C **13** Given that $n = 25, \sum x = 125, \sum x^2 = 650, \sum y = 100, \sum y^2 = 460$ and $\sum xy = 508$. Later on, it was found that two of the points $(8, 12)$ and $(6, 8)$ were wrongly entered as $(6, 14)$ and $(8, 6)$. Prove that $r = \frac{2}{3}$.

H **14** 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 **15** Two Judges in a beauty contest rank the 12 contestants as follows:

1st judge	1	2	3	4	5	6	7	8	9	10	11	12
2nd judge	12	9	6	10	3	5	4	7	8	2	11	1

What degree of agreement is there between the Judges?

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

C **16** 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 to discuss which pair of judges has nearest approach to beauty.

Answer: **2nd and 3rd judges has nearest approach**

$[\boldsymbol{\rho = 0.5515, 0.7333, 0.0545}]$

C **17** 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: $\boldsymbol{\rho = 0.8333}$

C **18** Calculate coefficient of correlation by spearman’s method from 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 **19** 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: $\boldsymbol{\rho = 0.5636}$

H **20** From the following data of the marks obtained by 8 students in Computer Networking (CN) and Compiler Design (CD) papers, Compute rank coefficient of correlation.

CN	15	20	28	12	40	60	20	80
CD	40	30	50	30	20	10	30	60

Answer: $\boldsymbol{\rho = 0.0298}$

H **21** The coefficient of rank correlation of marks obtained by 10 students in English and Economics was found to be 0.6. It was later discovered that the difference in ranks in the two subjects obtained by one of the students was wrongly taken as 7 instead of 1. Find the correct coefficient of rank correlation.

Answer: $\mathbf{0.3091}$

❖ REGRESSION ANALYSIS

- ✓ The regression analysis is concerned with the formulation and determination of algebraic expressions for the relationship between the two variables.
- ✓ We use the general form regression line for these algebraic expressions. The algebraic expressions of the regression lines are called 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.

❖ REGRESSION LINES

- ✓ Line of regression of y on x is $\mathbf{y - \bar{y} = b_{y\,x}\,(x - \bar{x})}$

$$b_{yx} = r\,\frac{\sigma_y}{\sigma_x} \quad \textbf{OR} \quad b_{yx} = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2} \quad \textbf{OR} \quad b_{yx} = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2}$$

- Line of regression of y on x is $\mathbf{x - \bar{x} = b_{x\,y}\,(y - \bar{y})}$

$$b_{xy} = r\,\frac{\sigma_x}{\sigma_y} \quad \textbf{OR} \quad b_{xy} = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum y^2 - (\sum y)^2} \quad \textbf{OR} \quad b_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(y-\bar{y})^2}$$

- Here, b_{xy} and b_{yx} are the regression coefficients & σ_x and σ_y are the standard deviation & $\bar{x}$ and $\bar{y}$ are the mean & r is the coefficient of correlation of x, y.

❖ PROPERTIES OF REGRESSION COEFFICIENTS

- The geometric mean (r) between two regression coefficients is given by $r = \sqrt{b_{yx} \times b_{xy}}$
- Both the regression coefficients will have the same sign. They are either both positive and both negative.
- The product of both b_{xy} and b_{yx} cannot be more than 1.
- The Sign of the coefficient of correlation is same as of the regression coefficients. It means,
  - If $r < 0$, then $b_{yx} < 0$ & $b_{xy} < 0$.
  - If $r > 0$, then $b_{yx} > 0$ & $b_{xy} > 0$.
- The arithmetic mean of the regression coefficients is greater than the correlation coefficient.

$$\frac{b_{xy} + b_{yx}}{2} > r$$

METHOD-7: EXAMPLES ON REGRESSION

H **1** Find the regression line of y on x for the following data:

x	2	3	4	4	5	6	6	7	7	8	10	10
y	1	3	2	4	4	4	6	4	6	7	9	10

Answer: $\mathbf{y = 0.99x - 0.92}$

C **2** Obtain two regression lines from the following data:

x	190	240	250	300	310	335	300
y	5	10	15	20	20	30	30

Answer: $\mathbf{x = 184.867 + 4.8533y}$, $\mathbf{y = -28.6233 + 0.1716x}$

<table>
<tr><td>H</td><td>3</td><td>Obtain two regression lines from the following data:
<table>
<tr><td>x</td><td>65</td><td>66</td><td>67</td><td>67</td><td>68</td><td>69</td><td>70</td><td>72</td></tr>
<tr><td>y</td><td>67</td><td>68</td><td>65</td><td>68</td><td>72</td><td>72</td><td>69</td><td>71</td></tr>
</table>
Answer: $\mathbf{x = 0.54x + 30.74,\ \ y = 0.665x + 23.78}$</td><td></td></tr>
<tr><td>H</td><td>4</td><td>Obtain the two lines of regression for the following data:
<table>
<tr><td>Sales (No. of tablets)</td><td>190</td><td>240</td><td>250</td><td>300</td><td>310</td><td>335</td><td>300</td></tr>
<tr><td>Advertising expenditure (Rs.)</td><td>5</td><td>10</td><td>12</td><td>20</td><td>20</td><td>30</td><td>30</td></tr>
</table>
Answer: $\mathbf{y = 0.1766x - 30.4221,\ \ x = 4.7357y + 189.0807}$</td><td>W-19
(7)</td></tr>
<tr><td>H</td><td>5</td><td>The amount of chemical compound (y), which were dissolved in 100 grams of water at various temperatures (x):
<table>
<tr><td>x</td><td>15</td><td>15</td><td>30</td><td>30</td><td>45</td><td>45</td><td>60</td><td>60</td></tr>
<tr><td>y</td><td>12</td><td>10</td><td>25</td><td>21</td><td>31</td><td>33</td><td>44</td><td>39</td></tr>
</table>
Find the equation of the regression line of y on x and estimate y if x = 50°C.
Answer: $\mathbf{y = 0.67x + 1.75,\ \ 35.25}$</td><td></td></tr>
<tr><td>C</td><td>6</td><td>A study of amount of rainfall and quantity of air pollution removed is:
<table>
<tr><td>Daily rainfall (0.01 cm)</td><td>4.3</td><td>4.5</td><td>5.9</td><td>5.6</td><td>6.1</td><td>5.2</td><td>3.8</td><td>2.1</td><td>7.5</td></tr>
<tr><td>Particulate removed unit</td><td>126</td><td>121</td><td>116</td><td>118</td><td>114</td><td>118</td><td>132</td><td>141</td><td>108</td></tr>
</table>
(a) Find the equation of the regression line to predict the particulate removed from the amount of daily rainfall.
(b) Find the amount of particulate removed when daily rainfall is 4.8 units.
Answer: $\mathbf{y = -6.336x + 153.24,\ \ 122.83}$</td><td></td></tr>
<tr><td>H</td><td>7</td><td>For following data Calculate the regression line of performing rating on experience and also estimate the probable performance if an operator has 11 years' experience.
<table>
<tr><td>Operator</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr>
<tr><td>Performance rating</td><td>78</td><td>36</td><td>98</td><td>25</td><td>75</td><td>82</td></tr>
<tr><td>Experience</td><td>84</td><td>51</td><td>91</td><td>60</td><td>68</td><td>62</td></tr>
</table>
Answer: $\mathbf{x = 11.428\,y - 29.38,\ \ 96.33}$</td><td></td></tr>
</table>

<table>
<tr><td>H</td><td>8</td><td>The following data regarding the height (y) and weight (x) of 100 students are given: $\sum x = 15000$, $\sum y = 6800$, $\sum x^2 = 2272500$, $\sum y^2 = 463025$, $\sum xy = 1022250$. Find the equation of regression line of height on weight.
Answer: $\mathbf{y = 0.1x + 53}$</td><td></td></tr>
<tr><td>C</td><td>9</td><td>The following values are available for the variable x & y. Obtain regression lines.
$n = 10, \sum x = 30, \sum y = 40, \sum x^2 = 222, \sum y^2 = 985, \sum xy = 384$.
Answer: $\mathbf{y = 2x - 2,\ \ x = 0.32y + 1.72}$</td><td></td></tr>
<tr><td>H</td><td>10</td><td>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: $\mathbf{y = 2.93x + 0.2568,\ \ y(1.5) = 4.6523,\ \ y(5.0) = 14.9083}$</td><td></td></tr>
<tr><td>C</td><td>11</td><td>The data for advertising and sale given below:
<table>
<tr><th></th><th>Adv. Exp. (x) (Rs lakh)</th><th>Sales (y) (Rs lakh)</th></tr>
<tr><td>Mean</td><td>10</td><td>90</td></tr>
<tr><td>Standard deviation</td><td>3</td><td>12</td></tr>
</table>
Correlation coefficient between prices is 0.8.
(a) Calculate the two regression lines.
(b) Find the likely sales when advertising expenditure is 15 lakhs.
(c) What should be the advertising expenditure if the company wants to attain a sales target of 120 lakhs?
Answer: $\mathbf{x = 0.2y - 8,\ \ y = 3.2x + 58,\ \ 106,\ \ 16}$</td><td></td></tr>
<tr><td>H</td><td>12</td><td>Find the regression lines from the following where $r = 0.5$:
<table>
<tr><th></th><th>x</th><th>y</th></tr>
<tr><td>Mean</td><td>60</td><td>67.5</td></tr>
<tr><td>Standard deviation</td><td>15</td><td>13.5</td></tr>
</table>
Answer: $\mathbf{y = 40.5 + 0.45x,\ \ x = 22.47 + 0.556y}$</td><td></td></tr>
</table>

H **13** Find the regression equation showing the capacity utilization on production from the following data: W-19 (7)

	Average	Standard deviation
Production (lakh units)	35.6	10.5
Capacity utilization (%)	84.8	8.5
Correlation coefficient	r = 0.62	

Estimate the production when capacity utilization is 70%.

Answer: $\mathbf{x = 0.7659y - 29.3483}$, $\mathbf{24.2647}$

H **14** A study of prices of a certain commodity at Hapur and Kanpur yields the below data:

	Hapur(Rs)	Kanpur(Rs)
Average price/kg	2.463	2.797
Standard deviation	0.326	0.207

Correlation coefficient between prices at Hapur and Kanpur is 0.774. Estimate the most likely price at Hapur corresponding to the price of 3.052 per kilo at Kanpur.

Answer: $\mathbf{2.774}$

PART-IV MISCELLANEOUS

METHOD-8: MISCELLANEOUS EXAMPLES

H **1** Find the mean, median, mode of the following frequency 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: $\mathbf{25.8472}$, $\mathbf{21.8478}$, $\mathbf{25.6579}$

C **2** Obtain the mean, mode and median for the following information:

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

Answer: $\mathbf{17.6}$, $\mathbf{16.6667}$, $\mathbf{17.2222}$

H **3** Obtain the mean, mode and median for the following information:

x	< 10	< 20	< 30	<40	<50	<60
f	12	30	57	77	94	100

Answer: **28, 25.625, 30.7407**

H **4** Find the average pocket expenses for the following data:

Pocket expenses (x)	18-21	22-25	26-35	36-45	46-55	56-65
Number of student (f)	8	3	55	36	20	12

Answer: **35.38**

C **5** Show that the median of following data is 31.7.

x	46-50	41-45	36-40	31-35	26-30	21-25	16-20	11-15
y	5	11	22	35	26	13	10	7

H **6** Find the mean and standard deviation of the following distribution:

Age	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Number of members	3	61	132	153	140	51	2

Answer: **54.72, 11.88**

C **7** An analysis of monthly wages paid to the workers of two firms A and B belonging to the same industry gives the following results:

	Firm A	Firm B
Number of workers	500	600
Average daily wage	186	175
Variance of distribution of wages	81	100

(a) Which firm has a larger wage bill?

(b) In which firm, is there greater variability in individual wages?

(c) Calculate average daily wages of all the workers in the firms A & B taken together.

Answer: **B, B, 180**

C **8** Lives of two models of refrigerators turned in for new models in a recent survey are given in the adjoining table.

Life (in year)	0-2	2-4	4-6	6-8	8-10	10-12
Model A	5	16	13	7	5	4
Model B	2	7	12	19	9	1

(a) What is the average life of each model of these refrigerators?

(b) Which model shows more uniformity?

Answer: **5.12, 6.16, B**

H **9** An analysis of monthly wages paid to the workers of two firms A and B belonging to the same industry gives the following results: W-19 (4)

	Firm A	Firm B
Number of workers	986	548
Average daily wage	52.5	47.5
Variance of distribution of wages	100	121

(a) Which firm has a larger wage bill?

(b) In which firm, is there greater variability in individual wages?

(c) Calculate average daily wages of all the workers in the firms A & B taken together.

Answer: **B, B, 49.87**

H **10** Prove that the skewness of the following data is 2.75.

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

H **11** Find the coefficient of variation, β_1 and β_2 for the following data:

x	170-180	180-190	190-200	200-210	210-220	220-230	230-240	240-250
y	52	68	85	92	100	95	70	28

Answer: **9.4, 0.003, 26.105**

C **12** Using moments method show that the following distribution is symmetric, Platykurtic.

x	0	1	2	3	4	5	6	7	8
f	1	8	28	56	70	56	28	8	1

<table>
<tr><td>C</td><td>13</td><td>In a partially destroyed laboratory record of an analysis of a correlation data following results are eligible:
variance=9, regression lines $8x - 10y + 66 = 0, 40x - 18y = 214$.
Find mean value of x & y, correlation coefficient between x & y, standard deviations.
Answer: $\mathbf{13,\ 17,\ 0.6,\ 3,\ 4}$</td><td></td></tr>
<tr><td>H</td><td>14</td><td>If coefficient of variance is 5, skewness is 0.5 and standard deviation is 2 then find the mean and mode of the distribution.
Answer: $\mathbf{40,\ 38}$</td><td></td></tr>
<tr><td>C</td><td>15</td><td>Find the mean and variance of the first n-natural numbers.
Answer: $\mathbf{\frac{n+1}{2},\ \frac{n^2-1}{12}}$</td><td></td></tr>
<tr><td>H</td><td>16</td><td>Find the mean, median and mode for the following frequency distribution:
<table>
<tr><td>x</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>f</td><td>4</td><td>7</td><td>8</td><td>10</td><td>6</td><td>6</td><td>4</td><td>2</td><td>2</td><td>1</td></tr>
</table>
Answer: $\mathbf{4.4,\ 4,\ 4}$</td><td></td></tr>
<tr><td>H</td><td>17</td><td>An insurance company obtained the following data for accident claims (in thousand rupees) from a particular region. Find its mean, median and mode.
<table>
<tr><td>Amount</td><td>1-3</td><td>3-5</td><td>5-7</td><td>7-9</td><td>9-11</td><td>11-13</td></tr>
<tr><td>Frequency</td><td>6</td><td>47</td><td>75</td><td>46</td><td>18</td><td>8</td></tr>
</table>
Answer: $\mathbf{6.47,\ 6.2533,\ 5.9825}$</td><td></td></tr>
</table>

★ ★ ★ ★ ★ ★ ★ ★ ★

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 types 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.
- ✓ 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. Then

$$z = \frac{t - E(t)}{S.E(t)}$$

is normally distributed with mean 0 and 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 $p_1 - P_2$	$\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$

❖ 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 in clear terms. (Always in Equality.)

- ✓ **Step 2:** Alternative hypothesis.
  - ➢ Set up H_1 , so that we could decide whether we should use one-tailed or two-tailed test. (Always less than or greater than or not equal).
- ✓ **Step 3:** Level of significance.
  - ➢ Select appropriate level of significance in advance depending on reality of estimates.
- ✓ **Step 4:** Critical region.
  - ➢ Given in data or find from statistical tabular table.
- ✓ **Step 5:** Test statistic.
  - ➢ Under null hypothesis compute the test statistic

$$z = \frac{t - E(t)}{S.E(t)}$$

- ✓ **Step 6:** 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.
  - ➢ It means if test statistic value belongs to critical Region, then we reject H_0 otherwise we accept H_0.

LARGE SAMPLE $(n \geq 30)$

❖ TEST FOR SINGLE PROPORTION

- ✓ This test is used to find the significant difference between proportion of sample & population.
- ✓ Let X be number of successes in n independent trials with constant probability P of success for each trial.
  - ➢ $E(X) = nP;\ V(X) = nPQ;\ Q = 1 - P =$ Probability of failure.
  - ➢ $E(p) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{nP}{n} = P$
  - ➢ $V(p) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}V(x) = \frac{1(PQ)}{n} = \frac{PQ}{n}$

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

$$\text{i.e. } z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

- ✓ This z is called test statistics which is used to test the significant difference of sample and population proportion.
- ✓ The probable limit for the observed proportion of successes is $p \pm z_\alpha\sqrt{\frac{PQ}{n}}$, 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: TEST 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.($z_{0.05} = 1.96$) **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%)($z_{0.05} = 1.645$) **Answer: The claim of manufacturer is not tenable(acceptable).**	

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? ($z_{0.05} = 1.96$) **Answer**: **The die is unbiased**.	
H	4	A coin was tossed 400 times and the head turned up 216 times. Test the hypothesis that the coin is unbiased. **Answer**: **The coin is unbiased**. ($\mathbf{z_{0.05} = 1.96}$)	

❖ TEST FOR DIFFERENCE BETWEEN PROPORTIONS

- ✓ 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

$$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \text{where } P = \frac{n_1p_1 + n_2p_2}{n_1 + n_2} \text{ and } Q = 1 - P.$$

METHOD – 2: TEST FOR DIFFERENCE BETWEEN PROPORTIONS

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? ($z_{0.05} < -1.645$) **Answer**: **The proportion of smokers is greater in city B than in A**.	
H	2	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 have concluded that the product of first factory is inferior than those of second? ($z_{0.05} > 1.645$) **Answer**: **Products do not differ in quality**.	

H	3	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 is a significant decrease in the consumption of tea after the increase in the excise duty? ($z_{0.05} > 2.33$) **Answer**: **There is significant decrease in consumption of tea.**	
C	4	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 of 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. ($z_{0.05} < -1.645$) **Answer**: **Proportion of correct answer by IP's is 15% more than those by AP's.**	

❖ TEST 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

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}, \text{where } \sigma \text{ is the standard deviation of the population.}$$

- ✓ 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: TEST 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 have an average life of 788 days, test the null hypothesis that $\mu \neq 800$ days at (a) 0.5 %, (b) 15% level of significance. ($z_{0.05} = 1.96$, $z_{0.15} = 1.44$) **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 weight 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. ($z_{0.05} = 1.96$) **Answer**: **The mean weight of the population is not 67 gms.**	
C	3	A college claims that its average class size is 35 students. A random sample of 64 students from 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. ($z_{0.05} > 1.645$) **Answer**: **The true mean class size is likely to be 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. ($z_{0.05} = 1.96$) **Answer: The machine does not require any adjustment.**	

❖ TEST FOR DIFFERENCE BETWEEN MEANS

- ✓ 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,

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

METHOD – 4: TEST FOR DIFFERENCE BETWEEN 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 lifetime of the two brands of light bulbs at a significance level of (a) 0.05, (b) 0.01? $(z_{0.05} - 1.96\,, z_{0.01} = 2.58)$ **Answer**: (**a**)**There is 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? $(z_{0.05} > 1.645\,, z_{0.01} > 2.33)$ **Answer**: (**a**)**The claim of company A is not valid**. (**b**)**The claim of company A is not valid**.	
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. $(z_{0.05} = 1.96)$ **Answer**: **No significant 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. $(z_{0.05} > 1.645)$

Type of wire	Mean resistance (ohms)	S.D. (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 supports the claim**.

❖ TEST FOR DIFFERENCE BETWEEN STANDARD DEVIATIONS

- ✓ 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 population standard deviation doesn't differ significantly, the static is

$$z = \frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}}, \text{ where } \sigma_1 \text{ and } \sigma_2 \text{ are population standard deviations.}$$

- ✓ When population standard deviations are not known then

$$z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}}, \text{ where } s_1 \text{ and } s_2 \text{ are sample standard deviations.}$$

METHOD – 5: TEST FOR DIFFERENCE BETWEEN STANDARD DEVIATIONS

C 1

Random samples drawn from two countries gave the following data relating to the heights of adult males:

	Country A	Country B
Standard deviation	2.58	2.5
Number in samples	1000	1200

Is the difference between the standard deviation significant? $(z_{0.05} = 1.96)$

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>n</th><th>S.D.</th></tr>
<tr><td>Girls</td><td>121</td><td>10</td></tr>
<tr><td>Boys</td><td>81</td><td>12</td></tr>
</table>
Is the difference between the standard deviations significant?
($z_{0.05} = 1.96$)
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>S.D.</td><td>34</td><td>28</td></tr>
</table>
Check whether the difference in the variability in yields is significant. ($z_{0.05} = 1.96$)
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 give a S.D. of 224 kg. Are the S. Ds significantly different? ($z_{0.05} = 1.96$)
Answer: The sample standard deviations are significantly different.</td><td></td></tr>
</table>

SMALL SAMPLE $(n < 30)$

❖ T-TEST FOR SINGLE 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

$$t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}, \quad \text{where } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

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

METHOD – 6: T-TEST FOR SINGLE 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 population having 56 as mean? (Value of t for 15 degrees of freedom at 5% level of significance is 2.131) **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 significant 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. (Value of t for 9 degree of freedom at 5% level of significance is 2.262) **Answer**: **There is no significant difference between population mean and sample mean.**	

H	4	The 9 items of a sample have the values 45,47,50,52,48,47,49,53,51. Does the mean of these values differ significantly from assumed mean 47.5? (Value of t for 9 degree of freedom at 5% level of significance is 2.262) **Answer: The mean of given values does not differ significantly from assumed mean 47. 5.**	
H	5	A manufacturer of external hard drives claims that only 10% of his drives require repairs within the warranty period of 12 months. If 5 of 20 of his drives required repairs within the first year, does this tend to support or refute the claim? **Answer: the claim should be refuted**	

❖ T-TEST FOR DIFFERENCE BETWEEN 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., $H_0 : \mu_1 = \mu_2$.
- ✓ 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 BETWEEN 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 significant 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 significant difference between two population means.</td><td>W-19
(3)</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.
(The value of t for 18 degree of freedom at 5% level is 1.734)
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 & 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>
Test whether the difference between the means of these two samples is significant. (The value of t for 9 degree of freedom at 5% level is 2.262)
Answer: **The average heat producing capacity of coal from two mines is not same**.</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>5</td>
<td>The following figures refer to observations in live independent samples:
<table>
<tr><td>Sample I</td><td>25</td><td>30</td><td>28</td><td>34</td><td>24</td><td>20</td><td>13</td><td>32</td><td>22</td><td>38</td></tr>
<tr><td>Sample II</td><td>40</td><td>34</td><td>22</td><td>20</td><td>31</td><td>40</td><td>30</td><td>23</td><td>36</td><td>17</td></tr>
</table>
Analyze whether the samples have been drawn from the population of equal means. [t at 5% level of significance for 18 d.f. is 2.1] Test whether the means of two populations are same at 5% level (t at 0.05=2.0739).
Answer: **Samples have been drawn from population with equal mean. Also, means of two populations are same**.</td>
<td>W-19
(7)</td>
</tr>
</table>

❖ 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

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \text{ with } v = n - 2 \text{ degrees of freedom.}$$

METHOD – 8: T-TEST FOR CORRELATION COEFFICIENT

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. (The value of t for 5 degree of freedom at 1% level is 4.032) **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 sample population? (The value of t for 13 degree of freedom at 5% level is 2.160) **Answer**: **The sample 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? (The value of t for 25 degree of freedom at 5% level is 2.06) **Answer**: **The sample 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? (The value of t for 623 degree of freedom at 5% level is 1.96) **Answer**: **It is highly significant**.	

❖ 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

$$F = \frac{(s_1)^2}{(s_2)^2}, \text{where } s_1^2 > s_2^2.$$

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 sample's 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. (F for 7 and 9 d.f. = 3.29)
Answer: There is no significant difference between the variances of two populations.</td><td></td></tr>
<tr><td>H</td><td>2</td><td>Two random samples reveal the following data:
<table>
<tr><th>Sample no.</th><th>Size</th><th>Variance</th></tr>
<tr><td>I</td><td>16</td><td>40</td></tr>
<tr><td>II</td><td>25</td><td>42</td></tr>
</table>
Test whether the samples come from the same normal population.
(F for 8 and 7 d.f. = 3.73)
Answer: The population variances are equal.</td><td></td></tr>
<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.
(F for 7 and 6 d.f. = 1.19)
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. (F for 5 and 6 d.f. = 4.39)
Answer: Samples have been drawn from the normal populations with same variance.</td><td></td></tr>
</table>

<table>
<tr>
<td>H</td>
<td>5</td>
<td>Two independent samples of 8 and 7 items respectively had the following values of the variable (weight in kg):
<table>
<tr><td>Sample I</td><td>9</td><td>11</td><td>13</td><td>11</td><td>15</td><td>9</td><td>12</td><td>14</td></tr>
<tr><td>Sample II</td><td>10</td><td>12</td><td>10</td><td>14</td><td>9</td><td>8</td><td>10</td><td>–</td></tr>
</table>
Do the two estimates of population variance differ significantly? Given that for (7,6) d.f. the value of F at 5% level of significance is 4.20 nearly.
Answer: **There is no significant difference between the variances of two population.**</td>
<td>W-19
(4)</td>
</tr>
<tr>
<td>H</td>
<td>6</td>
<td>Two samples of size 9 and 8 give the sum of squares of deviations from their respective means equal 160 inches and 91 inches respectively. Can they be regarded as drawn from two normal populations with the same variance? (F for 8 and 7 d.f. = 3.73).
Answer: **There is no significant difference between the variances of the population.**</td>
<td>W-19
(3)</td>
</tr>
</table>

❖ 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

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

- ✓ Note that the value of degree of freedom v for binomial, exponential and normal distribution is $n - 1$, $n - 2$ and $n - 3$, respectively.

METHOD–10: CHI-SQUARE TEST FOR GOODNESS OF FIT

<table>
<tr><td>C</td><td>1</td><td>Suppose that a die is tossed 120 times and the recorded data is as follows:
<table>
<tr><td>Face Observed(x)</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr>
<tr><td>Frequency</td><td>20</td><td>22</td><td>17</td><td>18</td><td>19</td><td>24</td></tr>
</table>
Test the hypothesis that the die is unbiased at $\alpha = 0.05$.
[χ^2 at 5% level of significance for 5 df is 11.070]
Answer: The die is unbiased.</td><td></td></tr>
<tr><td>H</td><td>2</td><td>The following table gives the number of accidents that took place in an industry during various days of the week. Test if accidents are uniformly distributed over the week.
<table>
<tr><td>Day</td><td>Mon</td><td>Tue</td><td>Wed</td><td>Thus</td><td>Fri</td><td>Sat</td></tr>
<tr><td>No. of accidents</td><td>14</td><td>18</td><td>12</td><td>11</td><td>15</td><td>14</td></tr>
</table>
[χ^2 at 5% level of significance for 5 df is 11.09]
Answer: The accidents are uniformly distributed over the week.</td><td></td></tr>
<tr><td>C</td><td>3</td><td>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).
<table>
<tr><td>(a)</td><td>1</td><td>5</td><td>20</td><td>28</td><td>42</td><td>22</td><td>15</td><td>5</td><td>2</td></tr>
<tr><td>(b)</td><td>1</td><td>6</td><td>18</td><td>25</td><td>40</td><td>25</td><td>18</td><td>6</td><td>1</td></tr>
</table>
Apply the χ^2-test of goodness of fit.
[χ^2 at 5% level of significance for 4 df is 9.488]
Answer: This normal distribution provides a good fit.</td><td></td></tr>
</table>

<table>
<tr><td>H</td><td>4</td><td>Suppose that during 400 five-minute intervals the air-traffic control of an airport received 0,1,2, … , 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 follows Poisson distribution with $\lambda = 4.6$.
[χ^2 at 5% level of significance for 8 df is 15.507]
Answer: Poisson distribution with $\lambda = 4.6$ provides a good fit.</td><td></td></tr>
<tr><td>C</td><td>5</td><td>Records taken of the number of male and female births in 830 families having four children are as follows:
<table>
<tr><td>No. of male births</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td></tr>
<tr><td>No. of female births</td><td>4</td><td>3</td><td>2</td><td>1</td><td>0</td></tr>
<tr><td>No. of families</td><td>32</td><td>178</td><td>290</td><td>236</td><td>94</td></tr>
</table>
Test whether data are consistent with hypothesis that the binomial law holds and the chance of male birth is equal to that of female birth, namely $p = q = \frac{1}{2}$. [χ^2 at 5% level of significance for 4 df is 9.49]
Answer: The data are not consistence with the hypothesis.</td><td>W-19
(7)</td></tr>
<tr><td>H</td><td>6</td><td>A die is thrown 276 times and the results of these throws are given below:
<table>
<tr><td>Number appeared on the die</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr>
<tr><td>Frequency</td><td>40</td><td>32</td><td>29</td><td>59</td><td>57</td><td>59</td></tr>
</table>
Test whether the die is biased or not.
[χ^2 at 5% level of significance for 5 df is 11.09]
Answer: The die is biased.</td><td>W-19
(4)</td></tr>
</table>

❖ 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 (E_{ij})
- H_1 : Attributes are dependent; that is, there is significant difference between observed frequencies (O_{ij}) and expected frequencies (E_{ij})

✓ The test static χ^2 for the analysis of $r \times c$ table is given by

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

✓ 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 following data of 80 persons.

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

[χ^2 at 5% level of significance for 2 df is 5.991]

Answer: **Hypertension and smoking habits are not independent**.

C **2** 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

[χ^2 at 5% level of significance for 4 df is 9.488]

Answer: **The hair color and gender are associated**.

H **3** 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 hypothesis that for an arbiter breakdown machine causing breakdown & the shift on which the breakdown occurs are independent.

[χ^2 at 5% level of significance for 4 df is 9.488]

Answer: **Machine causing breakdown and the shift are independent**.

★ ★ ★ ★ ★ ★ ★ ★ ★ ★

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. This 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 requiring to collect the data showing corresponding values of the variables under consideration.
- ✓ Suppose $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 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

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

❖ 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

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

- ✓ This will be minimum,

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

- ✓ Simplifying and expanding the above equations, we have

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

- ✓ Which implies

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

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

$$\sum y = an + b\sum x$$

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

✓ The normal equations for the best fitting straight line $y = ax + b$ is

$$\sum y = a\sum x + bn$$

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

METHOD – 1: FITTING A STRAIGHT LINE

C **1** By the method of least square, find the straight line that best fits the following data:

x	1	2	3	4	5
y	14	27	40	55	68

Answer: $\mathbf{y \ = 13.6x}$

H **2** Fit a straight line to the following data:

x	71	68	73	69	67	65	66	67
y	69	72	70	70	68	67	68	68

Answer: $\mathbf{y = 46.9394 + 0.3232x}$

H **3** The weight of a calf taken at weekly intervals are given below. Fit a straight line using method of least squares.

Age (x)	1	2	3	4	5	6	7	8	9	10
Weight (y)	52.5	58.7	65	70.2	75.4	81.1	87.2	95.5	102.5	108.4

Answer: $\mathbf{y = 45.6867 + 6.1752x}$

T **4** Fit a straight line for the given pairs of (x, y) which are $(1, 5)$, $(2, 7)$, $(3, 9)$, $(4, 10)$, $(5, 11)$.

Answer: $\mathbf{y = 3.9 + 1.5x}$

C **5** Fit a straight line $y = ax + b$ to the following data:

x	−2	−1	0	1	2
y	1	2	3	3	4

Answer: $\mathbf{y = 0.7x + 2.6}$

H **6** 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: $\mathbf{y = -0.5x + 8}$

H **7** 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:

P	13	18	23	27
W	51	75	102	119

Answer: $\mathbf{P = 0.2028W + 2.6580}$

H **8** The following show the gain in reading speed of 3 students in a speed-reading program, and the number of weeks they have been in the program:

No. of weeks	3	5	2	8	6	9	3	4
Speed gain	86	118	49	193	164	232	73	109

Find a straight line by the method of least squares.

Answer: $\mathbf{y = 3.3409 + 24.9318x}$

H **9** Fit a straight line for following data. Also, find y when $x = 2.8$.

x	2	5	6	9	11
y	2	4	6	9	10

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

C **10** By method of least squares, fit a linear relation of the form $P = a + bW$ to the following data, P is the pull required to lift a weight W. Also estimate P, when W is 150.

P	50	70	100	120
W	12	15	21	25

Answer: $\mathbf{P = -11.8005 + 5.3041W}$, $\mathbf{P(150) = 783.8145}$

❖ FITTING A PARABOLA

- ✓ 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. Let

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

- ✓ 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.

$$\sum y = na + b\sum x + c\sum x^2$$

$$\sum x\,y = a\sum x + b\sum x^2 + c\sum x^3$$

$$\sum x^2\,y = a\sum x^2 + b\sum x^3 + c\sum x^4$$

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

$$\sum y = a\sum x^2 + b\sum x + nc$$

$$\sum x\,y = a\sum x^3 + b\sum x^2 + c\sum x$$

$$\sum x^2\,y = a\sum x^4 + b\sum x^3 + c\sum x^2$$

METHOD – 2: FITTING A PARABOLA

C 1 Fit a second degree polynomial of y on x to the following data:

x	50	70	100	120
y	12	15	21	25

Answer: $\mathbf{y = 5.5259 + 0.1029x + 0.0005x^2}$

T **2** Fit a parabola to the following observations:

x	1	2	3	4	5	6
y	3.13	3.76	6.94	12.62	20.86	31.53

Answer: $\mathbf{y = 4.982 - 3.1199x + 1.2579x^2}$

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

x	1	2	3	5	6
y	1.1	5.8	17.5	55.9	86.7

Answer: $\mathbf{y = 2.7227 - 4.5528x + 3.0771x^2}$

H **4** Fit a parabola $y = a + bx + cx^2$ to the following data:

x	0	1	2	3	4
y	1	4	10	17	30

Answer: $\mathbf{y = 1.2 + 1.1x + 1.5x^2}$

H **5** 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	1.1	1.3	1.6	2.0	2.7	3.4	4.1

Answer: $\mathbf{y = 1.0357 - 0.1929\,x + 0.2429\,x^2}$

T **6** For 10 randomly selected observations, the following data were recorded.

Observation Number	1	2	3	4	5	6	7	8	9	10
Overtime Hours (x)	1	1	2	2	3	3	4	5	6	7
Additional units (y)	2	7	7	10	8	12	10	14	11	14

Determine the coefficient of regression using the non-linear form $y = a + b_1x + b_2x^2$.

Answer: $\mathbf{y = 1.8022 + 3.4823x - 0.2690x^2}$

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

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

x	−3	−2	−1	0	1	2	3
y	12	4	1	2	7	15	30

Answer: $\mathbf{y = 2.1190x^2 + 2.9286x + 1.6667}$

H **9** 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: $\mathbf{y = 10.4 - 11.0857x + 5.7143x^2, \;\; y(2.4) = 16.7087}$

C **10** Fit a relation of the form $R = a + bV + cV^2$ to the following data, where V is the velocity in km/hr. and R is the resistance in km/quintal. Estimate R when $V = 90$.

V	20	40	60	80	100	120
R	5.5	9.1	14.9	22.8	33.3	46.0

Answer: $\mathbf{R = 4.35 + 0.0024V + 0.0029V^2, \;\; R(90) = \; 28.0560}$

H **11** The following are the data on the drying time of a certain varnish and the amount of an additive that is intended to reduce the drying time? W-19 (7)

Amount of varnish additive(grams) ”x”	0	1	2	3	4	5	6	7	8
Drying time(hr.) “y”	12	10.5	10	8	7	8	7.5	8.5	9

I. Fit a second degree polynomial by the method of least square.
II. Use the result to predict the drying time of the varnish when 6.5 gms of the additive is being used.

Answer: $\mathbf{y = 12.1848 - 1.8465x + 0.1829x^2, \;\; y(6.5) = 7.9101}$

FITTING THE GENERAL CURVES

- **$y = ae^{bx}$**
  - Taking Logarithm on both sides $\log y = \log a + bx.$
  - Denoting $\log y = Y$ and $\log a = A$, we obtain $Y = A + bx.$
  - Find A and b using method of fitting a straight line.
  - Consequently $a = \text{Antilog}(A)$ can be calculated.
- **$y = ax^b$**
  - 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 and b using method of fitting a straight line.
  - Consequently $a = \text{Antilog}(A)$ can be calculated.
- **$y = ab^x$**
  - 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 and B using method of fitting a straight line.
  - Consequently $a = \text{Antilog}(A)$ and $b = \text{Antilog}(B)$ can be calculated.
- **$y = a + bx^2$**
  - Denoting $x^2 = X$, we obtain $y = a + bX.$
  - Find a and b using method of fitting a straight line.
- **$y = ax^2 + \frac{b}{x}$**
  - Multiplying by x both sides $yx = ax^3 + b.$
  - Denoting $xy = Y$ and $x^3 = X$, we obtain $Y = aX + b.$
  - Find a and b using method of fitting a straight line.
- **$pv^{\gamma} = C$**
  - $v = \left(\frac{C}{p}\right)^{\frac{1}{\gamma}} \Rightarrow v = C^{\frac{1}{\gamma}} p^{-\frac{1}{\gamma}}$

- Take logarithm both the sides $\log v = \frac{1}{\gamma}\log C - \frac{1}{\gamma}\log p$.
- Denoting $\log v = Y$, $\frac{1}{\gamma}\log C = A, -\frac{1}{\gamma} = B$ and $\log P = X$, we obtain $Y = A + BX$.
- Find A and B using method of fitting a straight line.
- Consequently $c = \text{Antilog}(\gamma A)$ and $\gamma = -\frac{1}{B}$ can be calculated.

METHOD – 3: FITTING THE GENERAL CURVES

<table>
<tr><td>C</td><td>1</td><td>Fit a curve of the best fit of the type $y = ae^{bx}$ to the following data:
<table>
<tr><td>x</td><td>1</td><td>5</td><td>7</td><td>9</td><td>12</td></tr>
<tr><td>y</td><td>10</td><td>15</td><td>12</td><td>15</td><td>21</td></tr>
</table>
Answer: $\mathbf{y = 9.4754 \cdot e^{(0.059)x}}$</td><td></td></tr>
<tr><td>H</td><td>2</td><td>Fit a curve of the best fit of the type $y = ae^{bx}$ to the following data:
<table>
<tr><td>x</td><td>1</td><td>2</td><td>3</td><td>4</td></tr>
<tr><td>y</td><td>1.65</td><td>2.7</td><td>4.5</td><td>7.35</td></tr>
</table>
Answer: $\mathbf{y = 1.0001 \cdot e^{(0.4993)x}}$</td><td></td></tr>
<tr><td>H</td><td>3</td><td>The population (p) of a small community on the outskirts of a city grows rapidly over a 20 −year period:
<table>
<tr><td>t</td><td>0</td><td>5</td><td>10</td><td>15</td><td>20</td></tr>
<tr><td>p</td><td>100</td><td>200</td><td>450</td><td>950</td><td>2000</td></tr>
</table>
As an engineer working for a utility company, you must forecast the population 5 years into the future in order to anticipate the demand for power. Employ an exponential model and linear regression to make this prediction.
Answer: $\mathbf{p = 97.915 \cdot e^{(0.151)t}}$</td><td></td></tr>
<tr><td>H</td><td>4</td><td>Fit a curve of the best fit of the type $y = ax^{b}$ to the following data:
<table>
<tr><td>x</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>y</td><td>27.8</td><td>62.1</td><td>110</td><td>161</td></tr>
</table>
Answer: $\mathbf{y = 7.3802 \cdot x^{1.9302}}$</td><td></td></tr>
</table>

C **5** Fit a curve of the best fit of the type $y = ax^b$ to the following data:

x	1	2	3	4	5
y	0.5	2	4.5	8	12.5

Answer: $\mathbf{y = 0.5x^2}$

C **6** Fit a curve of the best fit of the type $y = ab^x$ to the following data:

x	2	3	4	5	6
y	8.3	15.4	33.1	65.2	126.4

Answer: $\mathbf{y = 2.0495(1.9917)^x}$

H **7** Fit a curve of the best fit of the type $y = ab^x$ to the following data:

x	2	3	4	5	6
y	144	172.8	207.4	248.8	298.5

Answer: $\mathbf{y = 100.0230(1.2)^x}$

C **8** 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: $\mathbf{y = 4.1667 - 1.1111x^2}$

H **9** 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	3.88	7.66

Answer: $\mathbf{y = 0.5108x^2 - \frac{2.0826}{x}}$

H **10** The pressure P of the gas corresponding to various volume V is measured given by the following data, fit the data to the equation $PV^\gamma = C$.

P	50	60	70	80	90
V	64.7	51.3	40.5	25.9	78

Answer: $\mathbf{PV^{3.0931} = 11303240.36}$

* * *

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.

Seat No.: ________ Enrolment No.___________

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER– III (New) EXAMINATION – WINTER 2019

Subject Code: 3130006 **Date: 26/11/2019**
Subject Name: Probability and Statistics
Time: 02:30 PM TO 05:00 PM **Total Marks: 70**

Instructions:

1. **Attempt all questions.**
2. **Make suitable assumptions wherever necessary.**
3. **Figures to the right indicate full marks.**

Q.1 **(a)** In how many different ways can 4 of 15 laboratory assistants be chosen to assist with an experiment? **03**

(b) If 5 of 20 tires in storage are defective and 5 of them are randomly chosen for inspection (that is, each tire has the same chance of being selected), what is the probability that the two of the defective tires will be included? **04**

(c) The following are the data on the drying time of a certain varnish and the amount of an additive that is intended to reduce the drying time? **07**

Amount of varnish additive(grams)”x”	0	1	2	3	4	5	6	7	8
Drying time(hr) “y”	12.0	10.5	10.0	8.0	7.0	8.0	7.5	8.5	9.0

(i) Fit a second degree polynomial by the method of least square.
(ii) Use the result of (i) to predict the drying time of the varnish when 6.5 gms of the additive is being used.

Q.2 **(a)** If 3 balls are “randomly drawn” from a bowl containing 6 white and 5 black balls. What is the probability that one of the balls is white and the other two black? **03**

(b) The article “A Thin-Film Oxygen Uptake Test for the Evaluation of Automotive Crankcase Lubricants” reported the following data on oxidation-induction time (min) for various commercial oils: **04**
87, 103, 130, 160, 180, 195, 132, 145, 211, 105, 145, 153, 152, 138, 87, 99, 93, 119, 129
(i) Calculate the sample variance and standard deviation.
(ii) If the observations were re-expressed in hours, what would be the resulting values of the sample variance and sample standard deviation?

(c) In an examination, minimum 40 marks for passing and 75 marks for distinction are required. In this examination 45% students passed and 9% obtained distinction. Find average marks and standard deviation of this distribution of marks. **07**
[P(z=0.125)=0.05 and P(z=1.34)=0.41]

OR

(c) Distribution of height of 1000 students is normal with mean 165 cms and standard deviation 15 cms. How many soldiers are of height **07**
(i) less than 138 cms (ii) more than 198 cms (iii) between 138 and 198 cms.
[P(z=1.8)=0.4641, P(z=2.2)=0.4861]

Q.3 **(a)** Compute the coefficient of correlation between X and Y using the following data: **03**

X	2	4	5	6	8	11
Y	18	12	10	8	7	5

(b) An analysis of monthly wages paid to workers in two firms A and B belong to the same industry gave the following results. **04**

	Firm A	Firm B
No. of wages earners	986	548
Average monthly wages	Rs. 52.5	Rs. 47.5
Variance of distribution of wages	100	121

(a) Which firm pays out large amounts as wage bill?
(b) In which firm there is greater variability in individual wages?

(c) Obtain the two lines of regression for the following data: **07**

Sales (No. of tablets)	190	240	250	300	310	335	300
Advertising expenditure (Rs.)	5	10	15	20	20	30	30

OR

Q.3 **(a)** A sample of 20 items has mean 42 units and standard deviation 5 units. Test the hypothesis that it is a random sample from a normal population with mean 45 units.[t at 5% level for 19 d.f. is 2.09.] **03**

(b) A university warehouse has received a shipment of 25 printers, of which 10 are laser printers and 15 are inkjet models. If 6 of these 25 are selected at random to be checked by a particular technician, what is the probability that exactly 3 of those selected are laser printers (so that the other 3 are inkjets)? **04**

(c) Find the regression equation showing the capacity utilization on production from the following data: **07**

	Average	Standard deviation
Production (in lakh units)	35.6	10.5
Capacity utilization (in %)	84.8	8.5
Correlation coefficient	r = 0.62	

Estimate the production when capacity utilization is 70%.

Q.4 **(a)** Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant. Find the probability that in the next 18 samples, at least 4 samples contain the pollutant. **03**

(b) Goal scored by two teams A and B in a football season were as follows: **04**

No. of goals scored in a match	0	1	2	3	4
No. of matches played by team A	27	9	8	5	4
No. of matches played by team B	17	9	6	5	3

Find out which team is more consistent.

(c) Out of 800 families with 4 children each, how many families would be expected to have (i) 2 girls and 2 boys (ii) at least one boy (iii) no girl (iv) at most two girls? Assume equal probabilities for boys and girls. **07**

OR

Q.4 **(a)** Assume that the probability that a wafer contains a large particle of contamination is 0.01 and that the wafers are independent; that is, the probability that a wafer contains a large particle is not dependent on the characteristics of any of the other wafers. If 15 wafers are analyzed, what is the probability that no large particles are found? **03**

(b) 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 and 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? **04**

(c) If a publisher of nontechnical books takes great pains to ensure that its books are free of typographical errors, so that the probability of any given page containing at least one such error is .005 and errors are independent from page to page, what is the probability that one of its 400-page novels will contain (i) exactly one page with errors? (ii)At most three pages with errors? **07**

Q.5 **(a)** Samples of sizes 10 and 14 were taken from two normal populations with standard deviation 3.5 and 5.2. The sample means were found to be 20.3 and 18.6. Test whether the means of the two populations are the same at 5% level. [$t_{0.05}$=2.0739]. **03**

(b) Two independent samples of 8 and 7 items respectively had the following values of the variable (weight in kg): **04**

Sample I :	9	11	13	11	15	9	12	14
Sample II:	10	12	10	14	9	8	10	

Do the two estimates of population variance differ significantly? Given that for (7,6) d.f. the value of F at 5% level of significance is 4.20 nearly.

(c) Records taken of the number of male and female births in 830 families having four children are as follows: **07**

Number of male births	0	1	2	3	4
Number of female births	4	3	2	1	0
Number of families	32	178	290	236	94

Test whether the data are consistent with the hypothesis that the Binomial law holds and the chance of male birth is equal to that of female birth, namely p = q = 1/2. [χ^2 at 5% level of significance for 4 df is 9.49]

OR

Q.5 **(a)** Two samples of size 9 and 8 give the sum of squares of deviations from their respective means equal 160 inches and 91 inches square respectively. Can they be regarded as drawn from two normal populations with the same variance?
(F for 8 and 7 d.f. = 3.73). **03**

(b) A die is thrown 276 times and the results of these throws are given below: **04**

Number appeared on the die	1	2	3	4	5	6
Frequency	40	32	29	59	57	59

Test whether the die is biased or not. [χ^2 at 5% level of significance for 5 df is 11.09]

(c) The following figures refer to observations in live independent samples: **07**

Sample I:	25	30	28	34	24	20	13	32	22	38
Sample II:	40	34	22	20	31	40	30	23	36	17

Analyse whether the samples have been drawn from the population of equal means.
[t at 5% level of significance for 18 d.f is 2.1] Test whether the means of two population are same at 5% level (t at 0.05=2.0739)

****