20/12/16 Probability Module 1

Introduction:-

The theory of probability is one of the most useful and interesting branches of modern mathematics. It is (being) becoming prominent by its application in many fields of learning such as insurance, statistics, Biological Sciences, physical Sciences, Engineering etc.

Random Experiment:

If an Experiment is conducted, any number of times under essentially identical conditions, there is a set of all possible outcomes associated with it. If the result is not certain and is any one of the several possible outcomes, the experiment is called a random trail (or) an Random experiment. Outcomes are known as elementary events.

Sample Space:

Set of all possible outcomes of a random experiment is called a Sample space and is denoted by S.

Eg:- 1) In an experiment of tossing a coin possible outcomes are head and tail i.e $S = [H, T]$

2) In an experiment of throwing a dice possible outcomes are 1, 2, 3, 4, 5, 6, $S = \{1, 2, 3, 4, 5, 6\}$

Scanned with CamScanner

Event :

Any subset of a Sample space is called an Event and is denoted by E

Eg:- In an experiment of tossing a Coin, getting Head is an event (or) getting tail is an event.

$$E_1 = [H] \text{ or } E_2 = [T]$$

Mutually Exclusive Events :

Two events E_1, E_2 of a Sample Space S are Said to be mutually Exclusive if they have no Sample points in Common i.e $E_1 \cap E_2 = \phi$

Mutually Exclusive Events are Sometimes called as disjoint events.

Equally Likely Events :

Events are said to be equally likely when there is no reason to expect any one of them rather than any one of the others.

Eg:- when a Card is drawn from pack, any Card may be obtained. In this trail of all the 52 elementary events are equally likely.

Exhaustive Events :

All possible events in any trail is known as exhaustive events

Eg:- 1) In throwing a Coin there are 2 exhaustive elementary events i.e Head and tail

2) In Drawing 3 balls out of 9 balls in a box there are $^9C_3$ exhaustive Elementary Events.

Permutation:

A permutation of a number of objects or collection is the arrangement of objects in some definite Order.

i.e $$^nP_r = \frac{n!}{(n-r)!}$$

Combination:

The number of combination is an unordered Selection of an n distinct objects are collection taken r at a time is $c(n,r) = {}^nc_r = \frac{n!}{(n-r)!\,r!}$

21/12/16

Simple Event

An Event in a trail that cannot be further split is called a Simple event (or) An elementary event.

Classical Definition of probability.

In a Random Experiment, let there be n mutually exclusive and equally likely elementary events. Let E be an event of the experiment. If m elementary events from event E (are favourable to E), then the probability of E (probability of happening of E or chance of E), is defined as

$$P[E] = \frac{m}{n} = \frac{\text{number of elementary events in E}}{\text{Total number of elementary events in the random experiment.}}$$

note:

If $P(E) = 1$ the Event E is called certain event and

If $P(E) = 0$ the Event E is called an impossible event

1) Imp What is the probability for a Leap year to have 52 mondays and 53 Sundays.

sol: A Leap year has 366 days i.e 52 weeks 2 days. These 2 days can be any one of the following 7 ways

1 Monday Tuesday
2 Tuesday wednesday
3 wednesday Thursday
4 Thursday friday
5 friday Saturday
6 Saturday Sunday
7 Sunday Monday

Let E be the event of having 52 mondays and 53 Sundays in the year.

Total number of possible cases is $n = 7$

Number of favourable cases is $m = 1$

(Saturday and Sunday is the only favourable case)

$$\therefore P(E) = \frac{m}{n} = \frac{1}{7}$$

2) A class consists of 16 Boys and ~~10~~ 6 girls. If a Commity of 3 is choosen at random from the class, find the probability that

(i) 3 Boys are selected.

(ii) Exactly 2 girls are selected

Sol:- Total number of students = 16

$n(S)$ = no. of ways of choosing 3 from 16

$= {}^{16}C_3$ ways

(i) Suppose 3 boys are selected.

This can be done in ${}^{10}C_3$ ways

$n(E) = {}^{10}C_3$

$\therefore$ $P(E)$ = probability 3 boys are selected $= \dfrac{{}^{10}C_3}{{}^{16}C_3} = 0.214$

(ii) Suppose exactly 2 girls are selected then

$n(E) = {}^{6}C_2 \times {}^{10}C_1$

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

$$= \frac{{}^{6}C_2 \times {}^{10}C_1}{{}^{16}C_3}$$

$$= 0.2678$$

3) Two cards are selected at random from 10 cards numbered 1 to 10; find the probability that the Sum is even if

(i) 2 Cards are drawn together.

(ii) 2 Cards drawn one after other with replacement.

sol. (i) Suppose 2 Cards are drawn at a time.

$\therefore$ Number of ways of drawing 2 Cards at a time from 10 Cards $= 10_{C_2} = 45$ ways.

for the Sum on both the Cards to be even both the Cards should be even number or both the Cards Should be odd number.

2 Even number Cards Can be choosen from 5 even number cards $= 5_{C_2} = 10$ ways

$\therefore$ Total number of favourable outcomes $= 10+10 = 20$

$\therefore$ Required probability $= \frac{20}{45} = \frac{4}{9}$

(ii) Suppose the 2 Cards are choosen one after another with replacement. This can be done in $10 \times 10 = 100$ ways

for the Sum to be even both the Cards must be even or both the Cards must be odd.

Number of ways selecting 2 even Cards $= 5_{C_1} \times 5_{C_1}$
$= 25$ ways

Similarly

Number of ways of selecting 2 odd Cards $=$
$5_{C_1} \times 5_{C_1}$
$= 25$ ways

Required probability $= \frac{25+25}{100} = \frac{\cancel{50}\ 1}{\cancel{100}\ 2} = 1/2$

Scanned with CamScanner

HW

4) ✗ Two cards are selected at random from 10 each numbered 1 to 10. find the probability that the Sum is odd if

(i) if 2 cards are drawn

(ii) 2 Cards are drawn one after another with replacement.

(iii) 2 Cards are drawn one after another without replacement

Sol:- (i) 2 Cards can be drawn at a time from 10 cards in $^{10}C_2 = 45$ ways.

Let E_1 denote the event of 2 cards are such that the Sum is odd.

We must have one card even and another odd.

number of ways of doing it $= {}^5C_1 \times {}^5C_1 = 25$

Required probability $= \frac{25}{45} = \frac{5}{9}$

(ii) Let E_2 = The Sum is even when two cards are drawn one after another with replacement.

The no. of favourable cases = 50

∴ The no. of ways in which 2 Cards can be drawn one after another with replacement

$= {}^{10}C_1 \times {}^{10}C_1 = 100$

∴ Required probability $= \frac{50}{100} = 1/2$

(iii) The no. of favourable cases = 50

The number of cases that the two cards can be drawn one after another without replacement

$= {}^{10}C_1 \times {}^9C_1 = 90$.

∴ Required probability $= \frac{50}{90} = \frac{5}{9}$

24/12

5) 5 digit numbers are formed with 0,1,2,3,4 (not allowing a digit being repeated in any number. find the probability of getting 2 in the 10's place and 0 in the units place always.

Sol:- The total number of 5 digit no's using the digits 0,1,2,3,4 is $n = 5! - 4!$

$= 96$

$\underline{0}$ _ _ _ _ 5!

4! Zero (-)

Let E be the event of getting a number having 2 in 10's place and 0 in the units place so that number of favourable $= 3 \times 2 \times 1 \times 1 \times 1$

$= 6$

$\therefore P(E) = \frac{6}{96} = \frac{1}{16}$

6) Out of 15 items 4 are not in good Condition. 4 are selected at random. find the probability that

Imp

(i) All are not good

(ii) 2 are not good

2 are good.

$\frac{{}^{11}C_2}{{}^{15}C_4}$

Sol:

Total no. of items = 15

Number of ways of picking 4 items is ${}^{15}C_4$

Suppose 4 items are choosen which are not good

number of ways selecting $= {}^{4}C_4$

$\therefore$ probability all are not good $= \frac{{}^{4}C_4}{{}^{15}C_4}$

$= \frac{1}{1365}$

Exactly + Good, bad

${}^{11}C_2$

(ii) Suppose 2 items are not good.

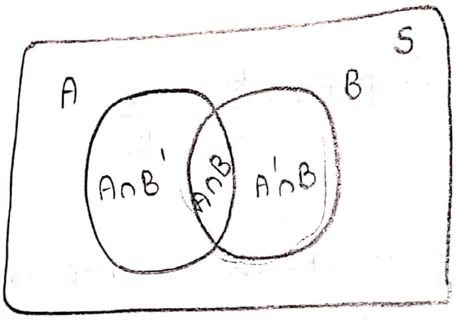
No. of ways of selecting of 2 bad items $= {}^4C_2 \times {}^{11}C_2$

$\therefore$ probability of getting 2 items $= \dfrac{{}^4C_2}{{}^{15}C_4}$

$= \dfrac{2}{455}$

Addition theorem of 2 Events -

For any 2 events A and B of a Sample space S prove that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



using the Venn diagram we observe that the events A and $A' \cap B$ are disjoint.

$$\therefore \quad A \cup B = A \cup (A' \cap B)$$

$$P(A \cup B) = P[A \cup (A' \cap B)]$$

$$= P(A) + P(A' \cap B)$$

Adding and Subtracting $P(A \cap B)$ RHS of above equation (form)

$$= P(A) + [P(A' \cap B) + P(A \cap B)] - P(A \cap B)$$

Since $A' \cap B$ and $A \cap B$ are disjoint events, the union of these 2 gives B.

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

1) A card is drawn from a well Sheffled pack of cards. what is the probability to have either a Spade and Or A's.

Sol:- Let S be a Sample space of all the Sample Events

$\therefore n(S) = 52$ (Total no. of cards)

Let A denote the event of getting a Spade and B denotes the event of getting an A's.

Then $A \cup B$ = Event of getting a spade or an A's

$A \cap B$ = Event of getting a spade and an A's.

$$\therefore P(A) = \frac{13}{52}$$

probability of A's $P(B) = \frac{4}{52}$

$$P(A \cap B) = \frac{1}{52}$$

By Addition theorem

$$P(A \cup B) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52}$$

$$= \frac{4}{13}$$

2) what is the probability that a card drawn at random from the pack of plane cards may be either a queen or a King.

sol: Let S be the Sample space associated with the drawing of a Card

$$\therefore n(S) = {}^{52}C_1 = 52$$

Let E_1 be the event of the card drawn being a queen

$$\therefore n(E_1) = {}^{4}C_1 = 4$$

Let E_2 be the event of the card drawn being a King

$$\therefore n(E_2) = {}^{4}C_1 = 4$$

But E_1, E_2 are mutually exclusive events.

Since $E_1 \cup E_2$ be the event drawing either a queen or a King.

we have $p(E_1 \cup E_2) = p(E_1) + p(E_2)$

$$= \frac{n(E_1)}{n(S)} + \frac{n(E_2)}{n(S)} \qquad p(E_1 \cap E_2) = \phi$$

$$= \frac{4}{52} + \frac{4}{52}$$

$$= \frac{2}{13}$$

3) Three students A, B, C are in running race; A & B have the same probability of winning and each is twice as likely to win as C. Find the probability that B or C wins.

Sol:- $S = (A \cup B \cup C)$ = sample space

By data $P(A) = 2P(C)$, $P(A) = P(B)$

We have $P(A) + P(B) + P(C) = 1$

$$2P(C) + 2P(C) + P(C) = 1$$

$$5P(C) = 1$$

$$P(C) = 1/5$$

$$P(A) = \frac{2}{5}, \quad P(B) = \frac{2}{5}$$

$\therefore P(A) + P(B) + P(C) = \frac{2}{5} + \frac{2}{5} + 1/5 = 1$

The probability of B or C wins

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

$$= \frac{2}{5} + \frac{1}{5} - 0$$

$$P(B \cup C) = \frac{3}{5}$$

Axioms of probability:-

i) Axiom of positivity :- $P(E) \geq 0$ for every subset $E(S)$

ii) Axiom of certainity :- $P(S) = 1$

iii) Axiom of union :- If E_1, E_2 are disjoint subsets of S. then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$

Note:-

→ Addition theorem of three events: Let A, B, C are any events of a sample space 's' then

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

Scanned with CamScanner

1) For a city three news papers A, B, C are being published. 'A' is read 20%, 'B' is read by 16%, 'C' is read by 14%, Both A & B are read by 18%. Both A & C are read by 5%, Both B & C are read by 4%. and all the three A, B, C are read by 2%. What is the percentage of the population that read atleast one paper.

Given $P(A) = \frac{20}{100}$ $\quad P(A\cap B) = \frac{8}{100}$

$P(B) = \frac{16}{100}$ $\quad P(B\cap C) = \frac{4}{100}$

$P(C) = \frac{14}{100}$ $\quad P(A\cap C) = \frac{5}{100}$

$P(A\cap B\cap C) = 2/100$

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

$$= \frac{20}{100} + \frac{16}{100} + \frac{14}{100} - \frac{8}{100} - \frac{4}{100} - \frac{5}{100} + \frac{2}{100}$$

$$= \frac{35}{100}$$

$\therefore P(A\cup B\cup C) = \frac{35\%}{100}$

$\therefore$ population that read atleast one paper is 35%.

<u>conditional event</u> :-

If E_1, E_2 are events of a sample spaces. And if E_2 occurs after the occurence of E_1, then the event of occurence of E_2 after the event E_1 is called conditional event of E_2 given E_1.

→ It is denoted by E_2/E_1.
Similarly we define E_1/E_2.

1) Example :- Two coins are tossed. The event of getting two tails given that there is atleast one tail is a conditional event.

2) Example :- Two unbiased dies are thrown. If the sum of the numbers thrown on them is '7'. The event of getting '1' on any one of them is a conditional event.

Conditional probability :- If E_1 & E_2 are two events in a sample space 's' and $P(E_1) \neq 0$, then the probability of E_2 after the event E_1 has occured is called the conditional probability of the event of E_2 given E_1 and is denoted by $P\left(\frac{E_2}{E_1}\right)$ (or) $P\left(\frac{E_1}{E_2}\right)$ and we define

$$P\left(\frac{E_2}{E_1}\right) = \frac{P(E_1 \cap E_2)}{P(E_1)}$$

Similarly we define

$$P\left(\frac{E_1}{E_2}\right) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

$$P\left(\frac{E_2}{E_1}\right) = \frac{n(E_1 \cap E_2)/n(S)}{n(E_1)/n(S)}$$

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

$$= \frac{\text{no. of elements in } (E_1 \cap E_2)}{\text{no. of elements in } E_1}$$

Multiplication Theorem of probability:-

statement:- If a random experiment if E_1, E_2 are two events such that $P(E_1) \neq 0$ and $P(E_2) \neq 0$ then $P(E_1 \cap E_2) = P(E_1) \cdot P(E_2/E_1)$

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

proof:-

Let 's' be the sample space associated with the random experiment. Let E_1, E_2 be two events of s. such that $P(E_1) \neq 0$, $P(E_2) \neq 0$ since $P(E_1) \neq 0$, by the definition of conditional probability of E_2 given E_1

$$P\left(\frac{E_2}{E_1}\right) = \frac{P(E_1 \cap E_2)}{P(E_1)}$$

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

since $P(E_2) \neq 0$ we have

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

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

Compound event :- When two or more events occur in conjunction with each other, their joint occurence is called compound event.

Example :- If two balls are drawn from a bag containing 4 green, 6 black, 7 white balls. The event of drawing two green balls. or two white balls is a compound event.

Independent and dependent events :-

If the occurence of the E_2 is not effected by the occurence or non occurence of the event E_1, then the event E_2 is said to be independent of E_1.

$$P(E_2/E_1) = P(E_2)$$

1) Two marbles are drawn in succession from a box containing 10 red, 30 white, 20 blue and 15 orange marbles with replacement being made after each draw. Find the probability that

i) Both are white.

ii) First is red and second is white.

Sol :- Total No. of marbles = 75

i) Let E_1 be the event of first drawn marble is white then $P(E_1) = \frac{30}{75} = \frac{30}{75}$

Let E_2 be the event of second drawn marble is also white then $P(E_2/E_1) = \frac{30}{75}$

∴ The probability of both marbles are white with replacement $P(E_1 \cap E_2) = P(E_1) \cdot P(E_2/E_1)$

$$= \frac{30}{75} \cdot \frac{30}{75}$$

$$= \frac{4}{25}$$

ii) Let E_1 be the event of the first drawn marble is red. Then $P(E_1) = \frac{10}{75}$

Let E_2 be the event of the second drawn marble is white then $P(E_2/E_1) = \frac{30}{75}$

∴ The probability that the first marble is red and the second marble is white.

$$P(E_1 \cap E_2) = P(E_1) \cdot P\left(\frac{E_2}{E_1}\right)$$

$$= \frac{10}{75} \cdot \frac{30}{75}$$

$$= \frac{4}{75} = \text{Ans}$$

2) Determine

i) $P(B/A)$ (or) B given A

ii) P(A) given B complement $P(A|B^c)$

Sol:- If A & B are events with

$P(A) = 1/3$, $P(B) = \frac{1}{4}$, $P(A \cup B) = 1/2$

$P(A) = 1/3$, $P(B) = 1/4$, $P(A \cup B) = \frac{1}{4}$

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

$$= \frac{P(A) + P(B) - P(A \cup B)}{P(A)}$$

$$P(B/A) = \frac{\cancel{1/12}\ \frac{1}{3}+\frac{1}{4}-\frac{1}{2}}{1/3}$$

$$= \frac{1/12}{1/3}$$

$$P(B/A) = \underline{1/4} \text{ ans}$$

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

$$P(A\cap B^c) = P(A) - P(A\cap B)$$

$$= 1/3 - \frac{1}{12} = 1/4$$

$$P(B^c) = 1 - P(B)$$

$$= 1 - 1/4 = 3/4$$

$$P(A/B^c) = \frac{1/4}{3/4} = \underline{1/3} \text{ ans}$$

3) A, B, C are aiming to shoot a balloon. 'A' will succeed four times out of 5 attempts. The chance of B to shoot the balloon is 3 out of 4. That of 'C' is 2 out of 3. If the three aim the balloon simultaneously, then find the probability that atleast two of them hit the balloon.

Sol:- $P(A) = \frac{4}{5}$

$P(B) = 3/4$

$P(C) = 2/3$

$\therefore$ The probability of A, B, C not hitting the target respectively or

$P(\bar{A}) = 1 - P(A) = 1 - \frac{4}{5} = 1/5$

$P(\bar{B}) = 1 - P(B) = 1 - \frac{3}{4} = 1/4$

$P(\bar{C}) = 1 - P(C) = 1 - 2/3 = 1/3$

Now the probablity that exactly 2 will hit the balloon is

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

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

$= \frac{4}{5} \cdot \frac{3}{4} \cdot 1/3 + \frac{4}{5} \cdot 1/4 \cdot 2/3 + 1/5 \cdot \frac{3}{4} \cdot 2/3$

$= \frac{13}{30}$

The probablity of all will hit the balloon.

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

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

$= \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \Rightarrow \frac{2}{5}$

The probablity that atleast two of them will hit the target

$= \frac{13}{30} + \frac{2}{5}$

$= \frac{5}{6}$

4) Three machines I, II, III produce 40%, 30%, 30% of the total number of items of factory. The pearcentages of defective items of this machines are 4%, 2%, 3%. If a one item is selected in a random. Find the probablity

that the item is deffective.

Sol:- Let A, B, C are be the events that the machines I, II & III be choosen respectively. and 'D' be the event which denotes the deffective item by data

$$P(A) = \frac{40}{100}$$
$$P(B) = 30/100$$
$$P(C) = 30/100$$

$$P(D/A) = 4/100$$
$$P(D/B) = 2/100$$
$$P(D/C) = 3/100$$

∴ The probability that the selected item at random is deffective is

$$P(D) = P(A) \cdot P(D/A) + P(B) \cdot P(D/B) + P(C) \cdot P(D/C)$$
$$= \left(\frac{40}{100}\right)\left(\frac{4}{100}\right) + \left(\frac{30}{100}\right)(2/100) + \left(\frac{30}{100}\right)(3/100)$$
$$= \frac{31}{1000}$$

28/12

5) 2 dice are thrown Let A be the event that the Sum of the points on the faces is 9. Let B be the event that atleast one number is 6. find

sol: (i) $P(A \cap B)$ (ii) $P(A \cup B)$ (iii) $P(A^c \cup B^c)$

There are 36 Simple Outcomes when 2 dice are thrown.

The event A = The Sum as 9 Occurs in the following way.

$$A = \{(3,6)(4,5)(5,4)(6,3)\}$$

$$P(A) = \frac{4}{36}$$

The event B that atleast 1 number is 6 occurs in the following way

$$B = \{(6,1)(6,2)(6,3)(6,4)(6,6)(1,6)(2,6)(3,6)(4,6)(5,6)\}$$

$$P(B) = \frac{11}{36}$$

now $A \cap B = \{(3,6)(6,3)\}$

(i) $P(A \cap B) = \frac{2}{36} = \frac{1}{18}$

(ii)

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

$$= \frac{4}{36} + \frac{11}{36} - \frac{1}{18}$$

$$= \frac{13}{36}$$

Scanned with CamScanner

(iii) $P(A^c \cup B^c) = P((A \cap B)^c)$

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

$$= 1 - \frac{1}{18}$$

$$= \frac{17}{18}$$

Scanned with CamScanner

*** State and prove BAYE'S theorem

V Imp

Statement:

$E_1, E_2, \ldots E_n$ are mutually exclusive and exhaustive events such that $P(E_i) > 0$ $(i = 1, 2, \ldots, n)$ in a Sample space S and A is any other event in S intersecting with every E_i (i.e A can only occur in Combination with any one of the events $E_1, E_2, \ldots, E_n$) such that $P(A) > 0$. If E_i is any of the events of $E_1, E_2, \ldots E_n$ where $P(E_1), P(E_2), \ldots, P(E_n)$ and $P(A/E_1), P(A/E_2), \ldots, P(A/E_n)$ are known then

$$P\left(E_K/A\right) = \frac{P(E_K)P(A/E_K)}{P(E_1)P\left(\frac{A}{E_1}\right) + P(E_2)P\left(\frac{A}{E_2}\right) + \cdots P(E_n)P\left(\frac{A}{E_n}\right)}$$

proof:

$E_1, E_2, \ldots, E_n$ are n events of S such that $P(E_i) > 0$ and $E_i \cap E_j = \phi$ for $i \neq j$ where $i, j = 1, 2, \ldots n$ also $E_1, E_2, \ldots; E_n$ are exhaustive events of S and A is any other event of S where $P(A) > 0$.

$$S = E_1 \cup E_2 \cup \cdots \cup E_n$$

$$A = A \cap S = A \cap (E_1 \cup E_2 \cup \cdots \cup E_n)$$

$$= (A \cap E_1) \cup (A \cap E_2) \cup \cdots \cup (A \cap E_n) \quad \text{--- (1)}$$

Here $(A \cap E_1), (A \cap E_2), \ldots (A \cap E_n)$ are mutually exclusive events then

$$P\left(E_K/A\right) = \frac{P(E_K \cap A)}{P(A)}$$

$$= \frac{P(E_K \cap A)}{P[(A \cap E_1) \cup (A \cap E_2) \cup \cdots \cup (A \cap E_n)]}$$

[from (1)]

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

By multiplication theorem / Conditional propobability

$$P\left(E_k/A\right) = \frac{P(E_k)\, P\left(A/E_k\right)}{P(E_1)\, P\left(A/E_1\right) + P(E_2)\, P\left(A/E_2\right) + \cdots + P(E_n)\, P\left(A/E_n\right)}$$

Sums:

1) In a certain College 25% of boys and 10% of girls are studying mathematics. The Girls Constitute 60% of the student body.

(a) what is the probability that mathematics is being studied.

(b) If a Student is selected at random and is found to be studying mathematics find the probability that the Student is girl.

(c) Student is boy.

sol:

Given $P(\text{Boys}) = \frac{40}{100} = \frac{2}{5}$

$P(\text{Girls}) = \frac{60}{100} = \frac{3}{5}$

probability that mathematics is studied given that the student is boy.

$\therefore P\left(M/B\right) = \frac{25}{100} = \frac{1}{4}$

probability that the mathematics is studied given that the student is a girl

$$\therefore P(M/G) = \frac{10}{100} = \frac{1}{10}$$

a) probability that the student studied mathematics

$$P(M) = P(B)\,P(M/B) + P(G)\,P(M/G)$$

$$= \frac{\cancel{2}}{5} \times \frac{1}{\cancel{4}_2} + \frac{3}{5} \times \frac{1}{10}$$

$$= \frac{1}{10} + \frac{3}{50}$$

$$= \frac{8}{50}$$

$$= \frac{4}{25}$$

b) By Bayes' theorem probability of mathematics student is a girl

$$P(G/M) = \frac{P(G)\,P(M/G)}{P(B)\,P(M/B) + P(G)\,P(M/G)}$$

$$= \frac{3/\cancel{5} \times 1/\cancel{10}\ 2}{4/\cancel{25}\ 5}$$

$$= \frac{3}{8}$$

c) By Bayes' theorem probability of mathematics student is a boy

$$P\left(B/M\right) = \frac{P(M)\,P\left(M/B\right)}{P(B)\,P\left(M/B\right) + P(G)\,P\left(M/G\right)}$$

$$\frac{\frac{\cancel{2}}{\cancel{5}}\cdot\frac{1}{\cancel{4}\,2}}{\frac{4}{\cancel{25}\,5}} = \frac{10}{16} = \frac{5}{8}$$

31/12

2) In a Bolt factory machines A, B, C manufacture 20%, 30% and 50% of the total of their output and 6%, 3%, and 2% are defectives. A bolt is drawn (from) at Random and found to be defective. find the probability that it is manufactured from (i) machine A (ii) machine B (iii) C.

Imp

Sol:- Let P(A), P(B), P(C) be the probabilities of the events that the bolts are manufactured by machine A, B, C respectively. Then $P(A) = \frac{20}{100}$

$P(B) = \frac{30}{100}$ and $P(C) = \frac{50}{100}$

Let D be the defective of the bolt, then

$$P\left(D/A\right) = \frac{6}{100},$$

$$P\left(D/B\right) = \frac{3}{100},$$

$$P\left(D/C\right) = \frac{2}{100}$$

(i) If a bolt is defective, then the probability it is from machine A

$$P\left(A/D\right) = \frac{P(A)\,P\left(D/A\right)}{P(A)\,P\left(D/A\right) + P(B)\,P\left(D/B\right) + P(C)\,P\left(D/C\right)}$$

$$= \frac{\frac{20}{100} \times \frac{6}{100}}{\frac{20}{100} \times \frac{6}{100} + \frac{30}{100} \times \frac{3}{100} + \frac{50}{100} \times \frac{2}{100}}$$

$$= \frac{12}{31}$$

(ii) If a bolt is defective, then the probability it is from machine B

$$P\left(B/D\right) = \frac{P(B)\,P\left(D/B\right)}{P(A)\,P\left(D/A\right) + P(B)\,P\left(D/B\right) + P(C)\,P\left(D/C\right)}$$

$$= \frac{\frac{30}{100} \times \frac{3}{100}}{\frac{20}{100} \times \frac{6}{100} + \frac{30}{100} \times \frac{3}{100} + \frac{50}{100} \times \frac{2}{100}}$$

$$= \frac{9}{31}$$

(iii) If a bolt is defective, then the probability it is from machine C

$$P\left(C/D\right) = \frac{P(C)\,P\left(D/B\right)}{P(A)\,P\left(D/A\right) + P(B)\,P\left(D/B\right) + P(C)\,P\left(D/C\right)}$$

$$= \frac{10}{31}$$

(3) A Businessman goes to hotels x, y, z 20%, 50%, 30% of the time respectively. It is known that 5%, 4%, 8% of the rooms in x, y, z Hotels have faulty plumbing D what is the probability that Business man's having faulty plumbing is assigned to hotel z.

Sol: Let $P(x), P(y), P(z)$ be the probabilities of the events of going to hotels x, y, z

$$P(x) = \frac{20}{100}, \quad P(y) = \frac{50}{100} \text{ and } P(z) = \frac{30}{100}$$

Let D be the faulty plumbing

(i) probability that Business man's having faulty plumbing is assigned to hotel z is

$$P\left(\frac{z}{D}\right) = \frac{P(z)\,P\left(\frac{D}{z}\right)}{P(x)\,P\left(\frac{D}{x}\right) + P(y)\,P\left(\frac{D}{y}\right) + P(z)\,P\left(\frac{D}{z}\right)}$$

$$= \frac{\frac{3\cancel{0}}{10\cancel{0}} \times \frac{8}{100}}{\frac{20}{100} \times \frac{5}{100} + \frac{50}{100} \times \frac{4}{100} + \frac{30}{100} \times \frac{8}{100}}$$

$$= \frac{4}{9}$$

(4) Of the 3 men, the chances that the politician, a businessmen or an acamecian will be appointed as Vice-chancellor of a university are 0·5, 0·3, 0·2 respectively. probability that research is promoted by these persons if they are appointed as VC are 0.3, 0.7, 0.8 respectively

(i) Determine the probability that research is promoted. 0·52

(ii) If Research is promoted what is the probability that VC is an academician 0·307

sol:- Let A, B, C be the events that a politician, businessmen or an academician will be appointed as V.C of the three men. Then

$P(A) = 0.5,$

$P(B) = 0.3$

$P(C) = 0.2$

The probabilities that research is promoted if they are appointed as V.C's are

$P\left(\frac{R}{A}\right) = 0.3,$

$P\left(R/B\right) = 0.7,$

and

$P\left(R/C\right) = 0.8$

(i) The probability that the research is promoted

$$= P(A) \cdot P\left(R/A\right) + P(B) \cdot P\left(R/B\right) + P(C) \cdot P\left(R/C\right)$$

$$= (0.5)(0.3) + (0.3)(0.7) + (0.2)(0.8)$$

$$= 0.52$$

(ii) The probability that research is promoted when the V.C is an academician

$$= P\left(C/R\right) = \frac{P(C) \cdot P\left(R/C\right)}{P(C) \cdot P\left(R/C\right) + P(B) \cdot P\left(R/B\right) + P(A) \cdot P\left(R/A\right)}$$

$$= \frac{0.16}{0.15 + 0.21 + 0.16}$$

$$= \frac{4}{13}$$

$$= 0\,30\,769$$

Boole's Inequality

If $A_1, A_2, \ldots A_n$ are n events then

(i) $P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$

(ii) $P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$

proof:-

(i) By Induction

$n=2$

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

probability of any event is ≥ 0 and ≤ 1

$A_1 \cup A_2$ is an event

$\therefore P(A_1 \cup A_2) \leq 1$ —— ①

$P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq 1$ —— ②

$1 + P(A_1 \cap A_2) \geq P(A_1) + P(A_2)$ —— ③

$\therefore P(A_1 \cap A_2) \geq P(A_1) + P(A_2) - 1$ —— ④

$\therefore$ The Statement is true for $n=2$

Assume that the statement is true for $n=r$

ie, $P\left(\bigcap_{i=1}^{r} A_i\right) \geq \sum_{i=1}^{r} P(A_i) - (r-1)$ —— ⑤

Then $n = r+1$

$P\left(\bigcap_{i=1}^{r+1} A_i\right) = P\left(\bigcap_{i=1}^{r} A_i \cap A_{r+1}\right)$ —— ⑥

$$P\left(\bigcap_{i=1}^{r+1} A_i\right) \geq P\left(\bigcap_{i=1}^{r} A_i\right) + P(A_{r+1}) - 1 \quad (\text{from } 4)$$

inequality we remove term

$$\geq \sum_{i=1}^{r} P(A_i) - (r-1) + P(A_{r+1}) - 1$$

$$\geq \sum_{i=1}^{r} P(A_i) + P(A_{r+1}) - r$$

$$\geq \sum_{i=1}^{r+1} P(A_i) - r$$

$$\therefore P\left(\bigcap_{i=1}^{r+1} A_i\right) \geq \sum_{i=1}^{r+1} P(A_i) - r$$

$\therefore$ The Induction of Hypothesis it is true for $n = r+1$

By the principle of mathematical Induction

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$$

(ii) $$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

proof :-

(i) By Induction

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - \underline{P(A_1 \cap A_2)} \quad \text{——} \; (1)$$

Quantities

$$P(A_1 \cap A_2) \geq 0$$

$n = 2$

$$\therefore P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$$

[$P(A_1 \cup A_2)$ is that Quantity which we get by subtracting a positive quantity from $P(A_1) + P(A_2)$]

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) \text{ only if } A_1 \cap A_2 = \phi$$

$$\therefore P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$$

This is true for $n = 2$.

Scanned with CamScanner

Suppose this is true for $n = r$

i.e., $P\left(\bigcup_{i=1}^{r} A_i\right) \leq \sum_{i=1}^{r} P(A_i)$ —— ③

Consider

$$P\left(\bigcup_{i=1}^{r+1} A_i\right) = P\left(\bigcup_{i=1}^{r} A_i \cup A_{r+1}\right)$$

$$\leq P\left(\bigcup_{i=1}^{r} A_i\right) + P(A_{r+1}) \quad \text{—— ④}$$

$$\leq \sum_{i=1}^{r} P(A_i) + P(A_{r+1}) \quad \text{—— ⑤}$$

$$\leq \sum_{i=1}^{r+1} P(A_i)$$

$$\therefore P\left(\bigcup_{i=1}^{r+1} A_i\right) \leq \sum_{i=1}^{r+1} P(A_i)$$

This is true for $n = r+1$.

$\therefore$ The Statement is true for $\forall n$ by principle of mathematical Induction.

Addition Theorem of N Events:
Or
Extension of General Law of Addition of probabilities

Statement:

For n events $A_1, A_2, \dots, A_n$ we have $P\left(\bigcup_{i=1}^{n} A_i\right) =$

$$\sum_{i=1}^{n} P(A_i) - \sum\sum_{1\le i \le j \le n} P(A_i \cap A_j) + \sum\sum\sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k) - \dots\dots + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n)$$

proof:-

For 2 events A_1 and A_2 we have Put $n=2$

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \quad \text{—— (1)}$$

Hence it is true for $n=2$

Let us now suppose that it is true for $n=r$

then $$P\left(\bigcup_{i=1}^{r} A_i\right) = \sum_{i=1}^{r} P(A_i) - \sum\sum_{1 \le i \le j \le r} P(A_i \cap A_j) + \dots\dots + (-1)^{r} P(A_1 \cap A_2 \cap \dots \cap A_r) \quad \text{—— (2)}$$

Now

$$P\left(\bigcup_{i=1}^{r+1} A_i\right) = P\left(\left(\bigcup_{i=1}^{r} A_i\right) \cup A_{r+1}\right)$$

$$= P\left(\bigcup_{i=1}^{r} A_i\right) + P(A_{r+1}) - P\left(\left(\bigcup_{i=1}^{r} A_r\right) \cap A_{r+1}\right)$$

[from (1)]

$$= P\left(\bigcup_{i=1}^{r} A_i\right) + P(A_{r+1}) - P\left(\bigcup_{i=1}^{r} (A_r \cap A_{r+1})\right)$$

(by distributive law)

$$= \sum_{i=1}^{r} (A_i) - \sum\sum_{1\le i<j\le r} P(A_i \cap A_j) + \cdots\cdots$$

$$+ (-1)^{r} P(A_1 \cap A_2 \cap \cdots \cap A_r) + P(A_{r+1}) - P\left(\bigcup_{i=1}^{r} (A_i \cap A_{r+1})\right) \quad [\text{from } ②]$$

now

$$P\left(\bigcup_{i=1}^{r+1} A_i\right) = \sum_{i=1}^{r} P(A_i) - \sum\sum_{1\le i<j\le r} P(A_i \cap A_j) + \cdots\cdots$$

$$\cdots + (-1)^{r} P(A_1 \cap A_2 \cap \cdots\cdots \cap A_m) + P(A_{r+1}) -$$

$$\left[\sum_{i=1}^{r} P(A_i \cap A_{r+1}) - \sum\sum_{1\le i<j\le r} P(A_i \cap A_j \cap A_{r+1}) + \cdots\cdots\right.$$

$$\left.\cdots + (-1)^{r-1} P(A_1 \cap A_2 \cap \cdots\cdots \cap A_m)\right] \quad [\text{from } ②]$$

$$= \sum_{i=r}^{r+1} P(A_i) - \left[\sum\sum_{1\le i\le j\le r} P(A_i \cap A_j) - \sum\sum_{1\le i<j\le r} P(A_i \cap A_j \cap A_{r+1})\right.$$

$$\left. + \cdots\cdots + (-1)^{r} P[(A_1 \cap A_2 \cap \cdots\cdots \cap A_{r+1})]\right]$$

$$= \sum_{i=1}^{r+1} P(A_i) - \sum\sum_{1\le i<j\le r+1} P(A_1 \cap A_2) + \cdots\cdots$$

$$\cdots + (-1)^{r+1} P[(A_1 \cap A_2 \cap \cdots \cap A_{r+1})]$$

It is true for $n = r$, it is also true for $n = r+1$ but we have proved in ① is true for $n = 2$

Hence by principle of mathematical induction it true for all +ve integer values of n.

Scanned with CamScanner

normal

Module 2

Random Variables:

Introduction:

Suppose S is the sample space of some experiment we Know that Outcomes of the experiment are the elements of the Sample space S and they need not be numbers. Sometimes (we wish) to assign a specify number to each Outcome.

Example: The no. of heads in tossing 2 Coins or 3 coins such assignment is called a Random Variable.

In the above example we may consider the Random Variable which is the number of heads

is.

Outcome :	HH	HT	TH	TT
Variables :	2	1	1	0

Random Variable:

2m

A Real Value x whose Value is determined by the Outcome of the Random Experiment is Called a Random Variable.

Types of Random Variable:

Random Variables is of 2 types:

(a) Discrete Random Variable.

(b) Continuous Random Variable.

(a) Discrete Random Variable:

A Random Variable X which can take only a finite number of discrete values in an interval of domain is called Discrete Random Variables

In other words,

If the Random Variables takes if the values only on the set $\{0, 1, 2, \dots, n\}$ is called a Discrete Random Variable.

Eg:- If a Coin is tossed $X(H) = 1$ if Head occurs.

$X(H) = 0$ if Tail Occurs.

(b) Continous Random Variable:

Random Variable X which can take Values Continously i.e which takes all possible Values in a given interval is called a Continous Random Variable.

Eg: The Height, age, weight, temperature

Probability Function of a Discrete Random Variable:

If for a discrete Random Variable X, the real Valued function $p(x)$ is such that $P(X)$ $P(X=x) = P(x)$ then $p(x)$ is called probability function or probability density function of a discrete Random Variable X properties of a probability function:

If $P(X)$ is a probability function of a Random Variable X then it is possess the following properties

(i) $P(X \geq 0)\ \forall\ X$

(ii) $\sum P(x) = 1$

(iii) $P(X)$ Cannot be negative for any Value of X

probability Distribution function:

Let X be a Random Variable. Then the probability distribution function associated with X is defined as the probability that the Outcome of an experiment will be one of the Outcomes for which $X(S) \leq x \rightarrow$ Variable

Sample Space

$$F_x(x) = P(X \leq x) : -\infty \leq x \leq \infty$$

Distribution function

is called the Distribution function of X:

properties of Distribution function:

(i) If F is the distribution function of a Random Variable X and if $a < b$ then

(a) $P(a < X < b) = F(b) = F(a)$

(b) $P(a \leq x \leq b) = P(x = a) + [F(b) - F(a)]$

(c) $P(a < x < b) = [F(b) - F(a)] - P(X = b)$

(d) $P(a \leq x \leq b) = [F(b) - F(a)] - P(X = b) + P(X = a)$

(2) $F(-\infty) = \underset{x \to -\infty}{Lt} F(x) = 0$

Discrete probability distribution: (probability mass function [PMF])

suppose X is a discrete Random Variable taking atmost infinite number of values $x_1, x_2, \dots x_n$ the probability of each possible outcome x_i, we associated a number $P_i = P(x = x_i) = P(x_i)$ is called the probability of x_i, $i = 1, 2, \dots\dots n$ must satisfy the following conditions

(i) $p(x_i \geq 0)$ for $\forall$ values of i

(ii) $\sum p(x_i) = 1$, $i = 1, 2, \dots\dots$ is called the PMF of Random Variable X and the set $p(x_i)$ is called Discrete probability distribution of the discrete Random Variable

Formulas:

Expectation, mean, Variance and Standard Variation of a discrete probability distribution:

(i) Expectation:

$$E(X) = \sum x\, p(x)$$

(2) Mean

$$\mu = \sum x\, p(x)$$

(3) Variance $V(X) = \sigma^2 = \sum (x_i - \mu)^2\, p(x)$

$$= E(x^2) - [E(x)]^2$$

(4) Standard deviation $= E(x^2) - \mu^2$

$$S.D = \sqrt{V(x)}$$

$$= \sqrt{\sigma^2}$$

$$= \sigma$$

Theorem 1:

Statement:

If X is a Random Variable, K is a Constant then

$E(X+K) = E(X) + K$

proof:-

By definition

$$E(X+K) = \sum_{i=1}^{n} (x+K)P(x)$$

$$= \sum_{i=1}^{n} x P(x) + \sum_{i=1}^{n} K P(x)$$

$$= E(x) + K \sum_{i=1}^{n} P(x)$$

$$E(X+K) = E(x) + K \quad \left[\because \sum P(x) = 1\right]$$

Hence proved.

Theorem 2:

Statement:

If X is a Random Variable, a, b are Constants that

$$E(aX+b) = a\,E(X) + b$$

proof:-

By definition

$$E(ax+b) = \sum_{i=1}^{n} (ax+b)\,P(x)$$

$$= \sum_{i=1}^{n} ax\,P(x) + \sum_{i=1}^{n} b\,P(x)$$

$$= a \sum_{i=1}^{n} x\,P(x) + b \sum_{i=1}^{n} P(x)$$

Statement:

*** If X and Y are 2 Random Variables then $E(X+Y) = E(X) + E(Y)$ provided $E(X)$ and $E(Y)$ exists.

vvv Imp (5M)

proof:

Let X Assume the values $x_1, x_2, \dots, x_n$ and Y assumes $y_1, y_2, \dots y_m$ then by definition

$$E(X) = \sum_{i=1}^{n} x_i P_i$$

$$E(Y) = \sum_{j=1}^{m} y_j P_j$$

Let $P_{ij} = (X = x_i \cap Y = y_j) = P(x_i, y_j)$

(This is called joint probability function of x and y)

The Sum $X+Y$ is also called a random Variable which can take $m \times n$ Values $(x_i + y_j)$, $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, m$.

$\therefore$ By definition

$$E(X+Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij}(x_i + y_j)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} x_i + \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} y_j$$

$$= \sum_{i=1}^{n} \left[x_i \sum_{j=1}^{m} P_{ij} \right] + \sum_{j=1}^{m} \left[y_j \sum_{i=1}^{n} P_{ij} \right]$$

$$= \sum_{i=1}^{n} x_i P_i + \sum_{j=1}^{m} y_j P_j = E(X) + E(Y)$$

note:

1) $E(X+Y+Z) = E(X) + E(Y) + E(Z)$.

2) $E(AX+BY) = A\,E(X) + B\,E(Y)$ where A and B are constants.

3) $E(X-\bar{X}) = 0$

Sums:

1) A Random Variable X has the following probability function

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

(i) Determine k

(ii) Evaluate $P(X<6)$, $P(X\geq 6)$, $P(0<X<5)$ and $P(0\leq X\leq 4)$

(iii) If $P(X\leq k > 1/2)$ find the minimum value of K

(iv) Determine the distribution function of X

(v) Mean

(vi) Variance.

Sol:- (i) we have $\sum_{i=0}^{7} p(x) = 1$

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

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

$$10k^2+10k-k-1=0$$

$$10k(k+1)-1(k+1)=0$$

$$(10k-1)(k+1)=0$$

$$k=1/10,\ k=-1$$

$$\therefore k=1/10 \quad \left(\because P(x)\geq 0 \text{ so } k\neq -1\right)$$

(ii) $P(x<6) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$

$= 0 + k + 2k + 2k + 3k + k^2$

$= k^2 + 8k$

$= \frac{1}{100} + \frac{8}{10}$

$= \frac{81}{100}$

$= 0.81$

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

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

$= \frac{19}{100}$

$P(X=6) + P(X=7)$

$2k^2 + 7k^2 + k$

$9k^2 + k$

$9\left(\frac{1}{100}\right) + \frac{1}{10} = \frac{19}{100}$

c) $P(0<X<5) = P(X=1) + P(X=2) + P(X=3) + P(X=4)$

$= k + 2k + 2k + 3k$

$= 8k$

$= \frac{8}{10}$

$= 0.8$

d) $P(0 \leq x \leq 4) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)$

$= 0 + k + 2k + 2k + 3k$

$= 8k$

$= \frac{8}{\ }$

(iii) The required minimum value of K is obtained as follows:-

$$P(X \leq 1) = P(X=0) + P(X=1)$$
$$= 0 + K$$
$$= \frac{1}{10}$$
$$= 0.1 \not> \frac{1}{2}$$

$$P(X \leq 2) = P(X \leq 1) + P(X=2)$$
$$= 0.1 + \frac{2}{10}$$
$$= 0.3 \not> \frac{1}{2}$$

$$P(X \leq 3) = P(X \leq 2) + P(X=3)$$
$$= 0.3 + \frac{2}{10}$$
$$= 0.5 \not> \frac{1}{2}$$

$$P(X \leq 4) = P(X \leq 3) + P(X=4)$$
$$= 0.5 + \frac{3}{10}$$
$$= 0.8 > 0.5 = \frac{1}{2}$$

$\therefore$ The minimum value of K for which $P(X \leq K > 1/2)$ is $K = 4$

(iv) The Distribution of function X is

X	$F(x) = P(X \le x)$
0	0
1	$k = 1/10$
2	$3k = 3/10$
3	$5k = 5/10$
4	$8k = 8/10$
5	$8k + k^2 = 81/100$
6	$81 + 3k^2 = 83/100$
7	$9k + 10k^2 = 1$

(v) Mean $= \sum_{i=0}^{7} x\, p(x)$

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

$$= k + 4k + 6k + 12k + 5k^2 + 12k^2 + 49k^2 + 7k$$

$$= 66k^2 + 30k$$

$$= 66\left(\frac{1}{100}\right) + 30\left(\frac{1}{10}\right)$$

$$= 0.66 + 3$$

$$E(X) = \mu = 3.66$$

(vi) Variance

$$v(x) = E(x^2) - (E(x))^2$$

$$= \sum x^2 p(x) - \mu^2$$

$$= \left[(0)^2(0) + (1)^2(k) + 2^2(2k) + 3^2(2k) + 4^2(3k) + 5^2(k^2) + 6^2(2k^2) + 7^2(7k^2+k)\right] - (3.66)^2$$

$$= \left[0 + 1(1/10) + 4(4/10) + 9\left(\frac{2}{10}\right) + 16\left(\frac{3}{10}\right) + 25\left(\frac{1}{100}\right) + 36\left(\frac{2}{100}\right) + 49\left(49\left(\frac{1}{10}\right)^2 + \frac{1}{10}\right)\right] - (3.66)^2$$

$$= 4.4 + 12.4 - (3.66)^2$$

$$v(x) = 3.4044 \neq \sigma^2$$

// Standard Variation $= \sqrt{v(x)}$

$$= \sqrt{3.4044}$$

$$= 1.845$$

HW

2) A Random Variable x has the following probability function

x	-3	-2	-1	0	1	2	3
$P(x)$	k	0.1	k	0.2	2k	0.4	2k

find (i) k (ii) mean (iii) Variance

Sol: (i) we have $\sum_{i=-3}^{3} P(x) = 1$

$\Rightarrow k + 0.1 + k + 0.2 + 2k + 0.4 + 2k = 1$

$6k + 0.7 = 1$

$6k = 1 - 0.7$ $\quad k = \frac{1-0.7}{6} = 0.05$

(ii) $\mu = \sum_{i=-3}^{3} x_i P(x_i)$

$= (-3)(k) + (-2)(k) + (-1)(k) + 0(0.2) + 1(2k) +$
$2(0.4) + 3(2k)$

but $k = 0.05$ $\quad 0.1$

$= (-3)(0.05) + (-2)(0.05) + (-1)(0.05) + 0(0.2) + 1(2(0.05))$
$+ 2(0.4) + 3(2(0.05))$

$E(x) = \mu = 0.8$

(iii) Variance

$v(x) = E(x2) - [E(x)]^2$

$= \sum x^2 p(x) - \mu^2$ $\quad \frac{143}{50}$

$= (-3)^2(0.05) + (-2)^2(0.05) + (-1)(0.05) + 0(0.2) +$
$1(2(0.05)) + 2^2(0.4) + 9(2(0.05))$

$= \frac{143}{50}$

$E(X-\mu)^2 = V(X)$

Theorem:

If x is a discrete random Variable then $v(ax+b) = a^2 v(x)$ where $v(x)$ is Variance of x and a,b are Constants.

proof:

Let $y = ax+b$ —— ①

Then $E(y) = E(ax+b)$

$= aE(x) + b$ —— ②

① - ② gives $(y - E(y)) = ax + \cancel{b} - aE(x) - \cancel{b}$

$(y - E(y)) = a(x - E(x))$

Squaring and taking Expectations on both Sides

$E((y - E(y))^2) = a^2 E[(x - E(x))^2]$

$v(y) = a^2 v(x)$ $\quad [\because E(x-\mu)^2 = v(x)]$

$v(ax+b) = a^2 v(x)$

case 1:

If $b=0$

then $v(ax) = a^2 x$

If $a=0$

then $v(b) = 0$

If $a=1$

then $v(x+b) = b(x)$

note: If x and y are 2 independent random Variables then $v(x \pm y) = v(x) \pm v(y)$

1) Let X denote the minimum of the 2 numbers that appear when a pair of fair dice is thrown once. Determine the

(i) Discrete probability distribution.

(ii) Expectation

(iii) Variance.

Sol: when 2 dice are thrown total number of Outcomes = 6 x 6

= 36

∴ Sample Space S is as follows

S = {(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)
(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)
(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)
(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)
(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)
(6,1) (6,2) (6,3) (6,4) (6,5) (6,6)}

If the random Variable X assignes the minim of its number in S then the Sample Space S is

$$S = \begin{Bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 & 3 & 3 \\ 1 & 2 & 3 & 4 & 4 & 4 \\ 1 & 2 & 3 & 4 & 5 & 5 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{Bmatrix}$$

The minimum number Could be {1, 2, 3, 4

for minimum 1, favourable cases are

$\{(1,1),(1,2),(1,3)(1,4)(1,5)(1,6)(2,1)(3,1)(4,1)(5,1)(6,1)\}$

// non repealed eg
$\{(2,1)$ but not
$(1,2)$ //

$\therefore P(X=1) = \frac{11}{36}$

Similarly $P(X=2) = \frac{9}{36}$

$P(X=3) = \frac{7}{36}$

$P(X=4) = \frac{5}{36}$

$P(X=5) = \frac{3}{36}$

$P(X=6) = 1/36$

i) The probability function is

x	1	2	3	4	5	6
$P(x)$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

(ii) $\sum_{i=1}^{6} E(X) = \sum_{i=1}^{6} x P(x)$

$= \frac{11}{36} + \frac{18}{36} + \frac{21}{36} + \frac{20}{36} + \frac{15}{36} + \frac{6}{36}$

$= \frac{91}{36}$

(iii) Variance

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

$= 1^2\left(\frac{11}{36}\right) + 2^2\left(\frac{9}{36}\right) + 3^2\left(\frac{7}{36}\right) + 4^2\left(\frac{5}{36}\right) + 5^2\left(\frac{3}{36}\right)$

$+ 6^2\left(\frac{1}{36}\right) - \left(\frac{91}{36}\right)^2$

$= \left(\frac{11}{36} + \frac{36}{36} + \frac{63}{36} + \frac{80}{36} + \frac{75}{36} + \frac{36}{36}\right) - \frac{91}{36}$

$= 5.83333$

2) find the mean and Variance of the uniform probability distribution given by $f(x) = \frac{1}{n}$ for $x =$ 2,3,....., n.

V imp

sol:- The probability distribution is

x	1	2	3	......	n
$f(x)$	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	......	$\frac{1}{n}$

(i) mean $= \mu = \Sigma x\, p(x)$

$= 1\left(\frac{1}{n}\right) + 2\left(\frac{1}{n}\right) + 3\left(\frac{1}{n}\right) + \ldots + n\left(\frac{1}{n}\right)$

$= \frac{1}{n}(1+2+3+\ldots+n)$

Scanned with CamScanner

$$= \frac{1}{\cancel{n}} \cdot \frac{\cancel{n}(n+1)}{2}$$

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

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

$$= 1^2 \cdot \frac{1}{n} + 2^2 \cdot \frac{1}{n} + \cdots\cdots + n^2 \frac{1}{n} - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{1}{n}\left(1^2 + 2^2 + \cdots\cdots + n^2\right) - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{1}{\cancel{n}}\left(\frac{\cancel{n}(n+1)(2n+1)}{6}\right) - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{n^2-1}{12}$$

10/1/17

3) A Sample of 4 items is selected at random from a box containing 12 items of which 5 are defective. find the expected number E of defective items.

Sol: Let x denote the number of defective items among 4 items drawn from 12 items.

Obviously x can take the values 0, 1, 2, 3 and 4

number of good items = 7

number of defective items = 5

$$P(x=0) = P(\text{no defective}) = \frac{{}^7C_4}{{}^{12}C_4} = \frac{7}{99}$$

$P(x=1) = P$ (one defective and 3 good items)

$$= \frac{{}^5C_1 \times {}^7C_3}{{}^{12}C_4} = \frac{35}{99}$$

$P(x=2) = P$ (Two defective and 2 good items)

$$\frac{{}^5C_2 \times {}^7C_2}{{}^{12}C_4} = \frac{42}{99}$$

$P(X=3) = P$ (3 defective and one good)

$$= \frac{{}^5C_3 \times {}^7C_1}{{}^{12}C_4} = \frac{14}{99}$$

$P(X=4) = P$ (all are defective)

$$= \frac{{}^5C_4}{{}^{12}C_4} = \frac{1}{99}$$

Discrete probability distribution is

x	0	1	2	3	4
$P(x)$	$\frac{7}{99}$	$\frac{35}{99}$	$\frac{42}{99}$	$\frac{14}{99}$	$\frac{1}{99}$

Expected no. of defective items $E(X) = \sum x\, p(x)$

$$0\left(\frac{7}{99}\right) + 1\left(\frac{35}{99}\right) + 2\left(\frac{42}{99}\right) + 3\left(\frac{14}{99}\right) + 4\left(\frac{1}{99}\right)$$

$$= \frac{5}{3}$$

4) A fair die is tossed. Let Random Variable X denote the twice the number appearing on the die

(i) write the probability distribution of X

(ii) mean

(iii) Variance

sol: Let x denote twice the number appearing on the face when a die is thrown.

Then x is a discrete random variable whose probability distribution is given by

(i)

$X = x_i$	2	4	6	8	10	12
$P(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

(ii) Mean $= E(x) = P_i x_i$

$$= 2 \times \frac{1}{6} + 4 \times \frac{1}{6} + 6 \times \frac{1}{6} + 8 \times \frac{1}{6} + 10 \times \frac{1}{6} + 12 \times \frac{1}{6}$$

$$= \frac{42}{6} = 7$$

now

$$E(x^2) = \Sigma x_i^2 P(x_i)$$

$$= 2^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6} + 8^2 \times \frac{1}{6} + 10^2 \times \frac{1}{6} + 12^2 \times \frac{1}{6}$$

$$= \frac{364}{6}$$

$$= 60.67$$

(iii) Variance = $E(x^2) - [E(x)]^2$

$= 60.67 - (7)^2$

$= 60.67 - 49$

$= 11.67$

Standard devariation = $\sqrt{11.67}$

$= 3.416138171$

Discrete Uniform distribution.

A Random Variable x has a discrete uniform distribution iff its probability distribution is given by $P(x) = 1/K$ for $x = x_1, x_2, \dots x_n$. Random Variable x is called discrete uniform distribution.

for eg:-

x	0	1
$P(x)$	$1/2$	$1/2$

Bernoulli's Distribution:

Random Variable x is [used] said to have a Bernoulli's Distribution with parameter p if its PMF is given by

$$P(X=x) = \begin{cases} p^x (1-p)^{1-x} & \text{for } x = 0,1 \\ 0 & \text{otherwise} \end{cases}$$

The parameter p satisfies $0 \le p \le 1$ & an $1-p$ is denoted by q

Binomial distribution:

A Random Variable X is said to follow a Binomial distribution if it assumes only non-negative values and its probability mass function is given b

$$P(X=x) = \begin{cases} {}^nC_x \, p^x q^{n-x} & x = 0,1,2,\dots\dots,n \text{ and } q = 1-p \\ 0 & \text{Otherwise} \end{cases}$$

Constants of Binomial distribution:

(i) Mean

Imp 2m / 5m

$$\mu = E(x) = \sum_{x=1}^{n} x\,p(x)$$

$$= \sum_{x=1}^{n} x \cdot {}^{n}C_{x}\, p^{x} q^{n-x}$$

$$= \sum_{x=1}^{n} \cancel{x} \cdot \frac{n}{\cancel{x}} \cdot {}^{n-1}C_{x-1}\, p^{x-1} \cdot p\, q^{n-x}$$

$$= np \sum_{x=0}^{n} {}^{n-1}C_{x-1}\, p^{x-1} q^{n-x}$$

$$= np\,(q+p)^{n-1} \quad \text{[from binomial distribution]}$$

$$\mu = np \quad [\because p+q=1]$$

(ii) Variance of Binomial distribution:

Consider $V(x) = E(x^2) - (E(x))^2$

$$E(x^2) = \sum_{x=0}^{n} x^2\, p(x)$$

$$= \sum_{x=0}^{n} x^2\, {}^{n}C_{x}\, p^{x} q^{n-x}$$

$$= \sum_{x=0}^{n} [x(x-1)+x]\, {}^{n}C_{x}\, p^{x} q^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1)\, {}^{n}C_{x}\, p^{x} q^{n-x} + \sum_{x=0}^{n} x\, {}^{n}C_{x}\, p^{x} q^{n-x}$$

$$= \sum_{x=0}^{n} \cancel{x}\cancel{(x-1)} \frac{n}{\cancel{x}} \cdot \frac{n-1}{\cancel{x-1}}\, {}^{n-2}C_{x-2}\, p^{x-2} p^{2} q^{n-x} + np$$

$$= \sum_{x=2}^{n} n(n-1)\, {}^{n-2}C_{x-2}\, p^{x-2}\, p^2\, q^{n-x} + np$$

$$= n(n-1)p^2 \sum_{x=0}^{n} {}^{n-2}C_{x-2}\, p^{x-2}\, q^{n-2} + np$$

$$= n(n-1)\, p^2 (q+p)^{n-2} + np$$

$$E(x^2) = n(n-1)\, p^2 + np$$

$$V(x) = n(n-1)p^2 + np - (np)^2$$

$$= \cancel{n^2p^2} - np^2 + np - \cancel{n^2p^2}$$

$$= np(1-p)$$

$$V(x) = npq$$

$$S.D = \sqrt{npq}$$

Moment Generating function of Binomial distribution:

$$M_X(t) = E(e^{tx})$$

$$= \sum_{x=0}^{n} e^{tx} \cdot p(x)$$

$$= \sum_{x=0}^{n} e^{tx}\, {}^nC_x\, p^x q^{n-x}$$

$$= \sum_{x=0}^{n} {}^nC_x\, (pe^t)^x\, q^{n-x}$$

$$M_X(t) = (q+pe^t)^n$$

Mode of Binomial distribution:

we know that

$$P(x) = {}^nC_x \, p^x q^{n-x} \quad \text{—— (1)}$$

$$P(x+1) = {}^nC_{x+1} \, p^{x+1} q^{n-(x+1)} \quad \text{—— (2)}$$

$$\text{eq (2)} \div \text{(1)} = \frac{{}^nC_{x+1} \, p^{x+1} q^{n-(x+1)}}{{}^nC_x \, p^x q^{n-x}} = \frac{P(x+1)}{P(x)}$$

$$P(x+1) = \frac{n-x}{x+1} \frac{p}{q} P(x)$$

problems:-

1) 10 Coins are tossed Simultaneously. find the probability of getting atleast

(i) 7 Heads.

(ii) 6 Heads.

Sol:- p = probability of getting head $= \frac{1}{2}$

q = probability of not getting a head $= \frac{1}{2}$.

The probability of getting x heads in a throw of 10 Coins is.

$$P(X=x) = P(x) = {}^{10}C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x} \qquad x = 0, 1, 2, \ldots, 10$$

(i) probability of getting atleast 7 Heads is given by $P(X \geq 7) = P(X=7) + P(X=8) + P(X=9) + P(X=10)$

$$= {}^{10}C_7 \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^{10-7} + {}^{10}C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^{10-8} +$$

Scanned with CamScanner

$$^{10}C_9\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)^{10-9}+{}^{10}C_{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^{10-10}$$

$$={}^{10}C_7\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^3+{}^{10}C_8\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2+{}^{10}C_9\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)+$$

$$^{10}C_{10}\left(\frac{1}{2}\right)^{10}$$

$$=\left(\frac{1}{2}\right)^{10}\left({}^{10}C_7+{}^{10}C_8+{}^{10}C_9+{}^{10}C_{10}\right)$$

$$=\frac{11}{64}$$

$$=0.1719$$

(ii) probability of getting atleast 6 Heads is given by $P(X\geq 6) = P(X=6)+P(X=7)+P(X=8)+P(X=9)+P(X=10)$

$$={}^{10}C_6\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^{10-6}+{}^{10}C_7\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^{10-7}+$$

$$={}^{10}C_8\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^{10-8}+{}^{10}C_9\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)^{10-9}+$$

$$^{10}C_{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^{10-10}$$

$$=0.3769$$

note:
exact $\Rightarrow 1$
atleast } $\geq$
atmost } $\leq$

Scanned with CamScann

2) If the probability of defective bolt is 0.2 (P) → Binomial / Poisson

(i) find mean

(ii) Standard deviation

for the distribution of bolts in a total of 400.

Sol: $n = 400$, $P = 0.2$, $q = 1 - P = 1 - 0.2 = 0.8$

(i) mean $= np$

$= (400)(0.2)$

$= 80$

(ii) Variance $= npq$

$= (400)(0.2)(0.8)$

$= 64$

Standard deviation $= \sqrt{npq}$

$= \sqrt{64}$

$= 8$

3) Out of 800 families with 5 children each, how many would you expect to have

VVV Imp

(a) 3 boys

(b) 5 Girls

(c) either 2 or 3 boys

(d) atleast one boy.

Assume equal probabilities for boys and girls

Sol: Let the no. of boys (in) each family is X.

Given: $P = \frac{1}{2}$ and $q = \frac{1}{2}$

$n = 5$

The probability distribution is

$$P(x) = {}^nC_x \, p^x \, q^{n-x}$$

$$= {}^5C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x}$$

(a) $P(3 \text{ Boys}) = P(X=3)$

$$= {}^5C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{5-3}$$

$$= {}^5C_3 \left(\frac{1}{2}\right)^5$$

$$= \frac{5}{16}$$

Thus for 800 families the probability of number of families having 3 boys $= \frac{5}{\cancel{16}} \times \cancel{800}^{50}$

$= 250$ families.

3 Boys ⟶ 250 families

(b) $P(5 \text{ Girls}) = P(X=0) = P(\text{no boys})$

$$= {}^5C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{5-0}$$

$$= \left(\frac{1}{2}\right)^5$$

$$= \frac{1}{32}$$

Thus for 800 families the probability of no. of families having 5 Girls $= \frac{1}{32} \times 800$

$= 25$ families.

(c) p(either 2 or 3 boys) = $p(x=2) + p(x=3)$

$$\Rightarrow {}^5c_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 + {}^5c_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2$$

$$\Rightarrow {}^5c_2 \left(\frac{1}{2}\right)^5 + {}^5c_3 \left(\frac{1}{2}\right)^5$$

$$\Rightarrow \frac{5}{\cancel{8}} \times \cancel{800}^{100}$$

$\Rightarrow$ 500 families.

(d) atleast one boy

$$p(x \geq 1) = 1 - p(x=0)$$

$$= \left(1 - \frac{1}{2^5}\right) 800$$

$$= 775$$

4) Mean and Variance of the Binomial distribution is 4 and 4/3 respectively. find $p(x \geq 1)$

sol:-

Given mean $= np = 4$ — ①

Variance $= npq = \frac{4}{3}$ — ②

÷ ② by ①

$$\frac{\cancel{np}q}{\cancel{np}} = \frac{\cancel{4}/3}{\cancel{4}} = \frac{1}{3}$$

$$q = 1/3$$

$$p = 1 - q$$

$$= 2/3$$

$np = 4$

$n\left(\frac{2}{3}\right) = \cancel{4}2$

$n = 6$

$P(X \geq 1) = 1 - P(X=0)$

$= 1 - 6c_0 \left(\frac{2}{3}\right)^0 \left(\frac{1}{3}\right)^{6-0}$

$= 1 - \left(\frac{1}{3}\right)^6$

$= \frac{728}{729}$

5) Determine the probability of getting the sum 6 exactly 3 times in 7th rows with a pair of fair dice.

Sol: In a single throw of a pair of fair dice a sum in 6 can occur in 5 ways i.e (1,5), (5,1), (2,4), (3,3), and (4,2)

Out of 36 ways.

Thus $p = \frac{5}{36}$ $q = \frac{31}{36}$ $(\because p+q=1)$

$n = 7$ [trails]

$\therefore$ probability of getting 6 exactly thrice in 7 throws $= 7c_3 \times p^3 q^{7-3}$

$= 7c_3 \left(\frac{5}{36}\right)^3 \left(\frac{31}{36}\right)^4$

$= 0.05155$

A coin is biassed in a way that a head is twice as likely to occur as a tail. If the coin is tossed 3 times find the probability of getting 2 tail and 1 Head.

Given $P(H) = 2P(T)$

we know that

$$P(H) + P(T) = 1$$

$$2P(T) + P(T) = 1$$

$$3P(T) = 1$$

$$P(T) = \frac{1}{3}$$

$$P(H) = \frac{2}{3}$$

Let getting a tail is a success and getting a head is a failure then $p = 1/3$ and $q = 2/3$.

Here $n = 3,\ x = 2$

$\therefore$ Required probability $= {}^3C_2 \left(1/3\right)^2 \left(2/3\right)^{3-2}$

$$= {}^3C_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^1$$

$$= \frac{2}{9}$$

Scanned with CamScann

Fitting a Binomial distribution.

Coin → $p, q = \frac{1}{2}$ in Binomial

Q) Fit a Binomial distribution to the following frequency distribution.

x	0	1	2	3	4	5	6
f	13	25	52	58	32	16	4

S: where n = number of trails

$= 6$

$N = \sum f = 200$

$\therefore$ Mean $= \frac{\sum fx}{\sum f} = \frac{0\times 13 + 1\times 25 + 2\times 52 + 3\times 58 + 4\times 32 + 5\times 16 + 6\times 4}{200}$

$= \frac{535}{200}$

$= 2.675$

$\therefore np = 2.675$

$6p = 2.675$

$p = 0.446$

$q = 1 - p$

$= 1 - 0.446$

$= 0.554$

fit a Binomial distribution

Scanned with CamScanner

x	f Observed frequency	$P(x) = {}^nC_x\, p^x q^{n-x}$ $= {}^6C_x (0.446)^x (0.554)^{6-x}$	Expected frequency $f(x) = NP(x)$ $= 200P(x)$
0	13	$P(0) = {}^6C_0 (0.446)^0 (0.554)^{6-0}$ $= 0.028$	$5.78 \cong 6$
1	25	$P(1) = {}^6C_1 (0.446)^1 (0.554)^{6-1}$ $= 0.139$	$27.9 \cong 28$
2	52	$P(2) = {}^6C_2 (0.446)^2 (0.554)^{6-2}$ $= 0.281$	$56.21 \cong 56$
3	58	$P(3) = {}^6C_3 (0.446)^3 (0.554)^{6-3}$ $= 0.301$	$60.3 \cong 60$
4	32	$P(4) = {}^6C_4 (0.446)^4 (0.554)^{6-4}$ $= 0.182$	$36.4 \cong 36$
5	16	$P(5) = {}^6C_5 (0.446)^5 (0.554)^{6-5}$ $= 0.058$	$11.74 \cong 12$
6	4	$P(6) = {}^6C_6 (0.446)^6 (0.554)^{6-6}$ $= 0.0079$	$1.57 \cong 2$
	200		200

HW

1) 4 coins are tossed 160 times. The no. of times x heads occur is given below: (N)

$n=4$

$p=q=1/2$

x	0	1	2	3	4
F	8	34	69	43	6

2) The probability that John hits the target is $1/2$. He fires 6 times. find the probability that he hits the target (a) Exactly 2 times

(b) more than 4 times

(c) Atleast once

$p = 1/2$

1) $P=\frac{1}{2}, q=\frac{1}{2}, n=4, N=160$

sol:-

x	P	$P(x) = {}^nC_x\, p^x q^{n-x}$	Expected frequency $f(x) = N\,P(x)$ $= 160\,P(x)$
		$p=q=1/2$	
0	8	$P(0) = {}^4C_0\, p^0 q^{4-0}$ $= (1/2)^4$	10
1	34	$P(1) = {}^4C_1\, p^1 q^{4-1}$ $= (1/4)$	40
2	69	$P(2) = {}^4C_2\, p^2 q^{4-2}$ $= (3/8)$	60
3	43	$P(3) = {}^4C_3\, p^3 q^{4-3}$ $= (1/4)$	40
4	6	$P(4) = {}^4C_4\, p^4 q^{4-4}$ $= (1/16)$	10
	160		160

2.

sol:- probability of hitting a target $= p = \frac{1}{2}$

probability of no hit $= q = \frac{1}{2}$

number of trials $= n = 6$

number of hits (successes) $= X$

(i) $P(\text{exactly } 2 \text{ times}) = P(X=2)$

$= 6_{C_2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^4$

$= \frac{15}{2^6}$

$= 0.234$

(ii) $P(\text{more than } 4 \text{ times}) = P(X>4)$

$= P(X=5) + P(X=6)$

$= 6_{C_5} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right) + 6_{C_6} \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^0$

$= \frac{6}{2^6} + \frac{1}{2^6}$

$= \frac{7}{2^6}$

$= 0.1094$

(iii) $P(\text{atleast once}) = P(X \geq 1) = 1 - P(X=0)$

$= 1 - 6_{C_0} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^6$

$= 1 - \frac{1}{2^6} = 0.9844$

11/1/17

Poisson Distribution

A Random Variable X is said to follow a Poisson Distribution if it assumes only non-negative values and its probability mass function is given by

$$P(x, \lambda) = P(X=x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & x = 0, 1, 2, \ldots, \lambda > 0 \\ 0 & \text{otherwise} \end{cases}$$

Here $\lambda > 0$ is called the parameter of the distribution

note:-

1) $$\sum_{x=0}^{\infty} P(X=x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^{x}}{x!}$$

$$= e^{-\lambda} \cdot e^{\lambda}$$

by e^{α} expansion.

$$= 1$$

This is known as probability function.

2) The distribution function is $F(x) = P(X \leq x) = \sum_{r=0}^{x} P(r)$

$$= \sum_{r=0}^{x} \frac{e^{-\lambda}\lambda^{r}}{r!}$$

$$= e^{-\lambda} \sum_{r=0}^{x} \frac{\lambda^{r}}{r!} \quad , r = 0, 1, 2, \ldots$$

mean = Variance

Constants of poisson distribution :-

(i) Mean

$$\mu = E(x) = \sum_{x=0}^{\infty} x\, p(x)$$

$$= \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda} \lambda^{x}}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \cancel{x} \cdot \frac{\lambda \cdot \lambda^{x-1}}{\cancel{x}(x-1)!}$$

$$= \lambda e^{-\lambda} \cdot \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$\mu = \lambda e^{-\lambda} \left[1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots\cdots \right]$$

$$= \lambda \cancel{e^{-\lambda}} \cancel{e^{\lambda}} \left[\because e^{x} = 1 + x + \frac{x^2}{2!} + \cdots\cdots\cdots \right]$$

$$= \lambda$$

$\therefore$ Mean of poisson distribution is λ.

(ii) Variance

$$V(x) = E(x^2) - (E(x))^2$$

consider $E(x^2) = \sum_{x=0}^{\infty} x^2 p(x)$

$$= \sum_{x=0}^{\infty} (x(x-1) + x) \frac{e^{-\lambda} \lambda^{x}}{x!}$$

$$= \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda} \lambda^{x}}{x!} + \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^{x}}{x!}$$

$$= \sum_{x=0}^{\infty} \frac{x(x-1)e^{-\lambda}\lambda^{x-2}\cdot\lambda^2}{x(x-1)(x-2)!} + \lambda$$

$$= e^{-\lambda}\lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda$$

$$= e^{-\lambda}\lambda^2\left[1+\lambda+\frac{\lambda^2}{2!}+\ldots\ldots\right] + \lambda$$

$$= \cancel{e^{-\lambda}}\lambda^2 \cancel{e^{\lambda}} + \lambda$$

$$E(x^2) = \lambda^2 + \lambda$$

$$V(x) = \cancel{\lambda^2} + \lambda - \cancel{\lambda^2}$$

$$= \lambda$$

$\therefore$ Mean = Variance = λ

Standard deviation of poisson distribution = $\sqrt{\lambda}$

Recurrence Relation for poisson distribution

Imp

we have

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$P(x+1) = \frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!}$$

$$P(x+1) = \frac{e^{-\lambda}\lambda^x\cdot\lambda}{(x+1)x!}$$

$$= \frac{e^{-\lambda}\cdot\lambda^x}{x!}\cdot\frac{\lambda}{x+1}$$

$$P(x+1) = P(x)\cdot\frac{\lambda}{x+1} \Rightarrow P(x+1) = \frac{\lambda}{x+1}P(x)$$

Moment Generating function of poisson distribution.

$$m_x(t) = \sum_{x=0}^{\infty} e^{-tx} p(t)$$

$$= \sum_{x=0}^{\infty} e^{-tx} \frac{e^{-\lambda} \lambda^x}{x!}$$

$$= \sum_{x=0}^{\infty} e^{-\lambda} \frac{(e^{t} \cdot \lambda)^x}{x!}$$

$$= e^{-\lambda} \left[1 + \frac{\lambda e^t}{1!} + \frac{(\lambda e^t)^2}{2!} + \dots \right]$$

$$= e^{-\lambda} e^{\lambda e^t}$$

$$m_x(t) = e^{\lambda(e^t - 1)}$$

problems:

1) If the probability that an individual suffers a bad reaction from a certain injection is 0.001. Determine the probability that out of 2000 individual

(i) Exactly 3.

(ii) more than 2 individuals.

(iii) none

(iv) more than one individual suffered by 1 individual.

Binomial / Poisson

sol:

$P = 0.001$, $n = 2000$.

$$\lambda = np$$

$$= (0.001)(2000)$$

$$\lambda = 2$$

Scanned with CamScann

(i) $P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

$$P(3) = P(X=3) = \frac{e^{-2}2^{3}}{3!}$$

$$= 0.1804$$

(ii) $P(\text{more than } 2) = P(X \geq 2)$

$$= 1-[P[x=0) + P(x=1) + P(x=2)]$$

$$= 1-\left[\frac{e^{-2}2^{0}}{0!} + \frac{e^{-2}2^{1}}{1!} + \frac{e^{-2}2^{2}}{2!}\right]$$

$$= 1-e^{-2}[1+2+2]$$

$$= 1-e^{-2}[5]$$

$$= 0.3233$$

(iii) $P(\text{none}) = P(X=0) = \frac{e^{-2}2^{0}}{0!}$

$$= 0.1353$$

(iv) $P(\text{more than one}) = P(X>1)$

$$= 1-[P(X=0)+P(X=1)]$$

$$= 1-\left[\frac{e^{-2}2^{0}}{0!} + \frac{e^{-2}2^{1}}{1!}\right]$$

$$= 1-e^{-2}[1+2]$$

$$= 1-3e^{-2}$$

$$= 0.594$$

2) If a random Variable has a poisson distribution such that $P(1) = P(2)$ find (i) mean of the distribution (ii) $P(4)$ (iii) $P(X \geq 1)$ (iv) $P(1 < x < 4)$

sol:-

Given $P(1) = P(2)$

$$\frac{\cancel{e^{-\lambda}} \cancel{\lambda^1}}{1!} = \frac{\cancel{e^{-\lambda}} \lambda^{\cancel{2}}}{2!}$$

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

$$\lambda = 2$$

(i) mean = 2

(ii) $P(4) = \dfrac{e^{-2} 2^4}{4!}$

$= 0.0902$

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

$= 1 - P(X = 0)$

$= 1 - \dfrac{e^{-2} 2^0}{0!}$

$= 1 - e^{-2}$

$= 0.8647$

(iv) $P(1 < x < 4) = P(x = 2) + P(x = 3)$

$= \dfrac{e^{-2} 2^2}{2!} + \dfrac{e^{-2} 2^3}{3!}$

$= 0.4511$

17/1/14

3) using recurrence formula find the probabilities when $x = 0, 1, 2, 3, 4$ and 5. If the mean of poisson distribution is 3

sol:- Given mean of the poisson distribution is 3 i.e

$\lambda = 3$

Now the poisson distribution is $P(x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$

$$= \frac{e^{-3}3^{x}}{x!}$$

$$\therefore P(x=0) = \frac{e^{-3}3^{0}}{0!} = e^{-3} = 0.0498$$

Cannot apply recurrence formula $\therefore$ %

using recurrence formula

$$P(x+1) = \frac{\lambda}{x+1} P(x) \quad \text{——— (1)}$$

put $x=0$ in (1)

$$P(1) = \frac{3}{1} P(0)$$

$$= 3(0.0498)$$

$$P(1) = 0.1494$$

put $x = 2$ in (1), we have

$$P(2) = \frac{3}{1+1} P(1)$$

$$= \frac{3}{2}(0.1494)$$

$$= 0.2241$$

$$= 0.2241$$

2)

put $x=2$ in ①, we have

$$P(3)=\frac{3}{2+1}P(2)$$

$$=0.2241$$

put $x=3$ in ①

$$P(4)=\frac{3}{3+1}P(3)$$

$$=\frac{3}{4}(0.2241)$$

$$=0.1681$$

put $x=4$ in ①

$$P(5)=\frac{3}{4+1}P(4)$$

$$=\frac{3}{5}(0.1681)$$

$$=0.1009$$

4) Suppose 2% of the people on the $\underline{\overset{\lambda}{\text{average}}}$ are lefthanded. find (i) the probability of finding 3 are more left handed

(ii) The probability of finding none are 1 lefthand

sol:- Let X be the number of left handed

Given mean $=\lambda=2\%=0.02$

we have $P(x)=\dfrac{e^{-\lambda}\lambda^{x}}{x!}$

$$=\frac{e^{-0.02}(0.02)^{x}}{x!}$$

(i) $P(X \geq 3) = 1 - P(X<3)$

$= 1 - [P(X=0) + P(X=1) + P(X=2)]$

$= 1 - \left[\frac{e^{-0.02}(0.02)^0}{0!} + \frac{e^{-0.02}(0.02)^1}{1!} + \frac{e^{-0.02}(0.02)^2}{2!}\right]$

$= 1.3077 \times 10^{-6}$

(ii) $P(X \leq 1) = P(0) + P(1)$

$= \frac{e^{-0.02}(0.02)^0}{0!} + \frac{e^{-0.02}(0.02)^1}{1!}$

$= 0.999$

5) If x is a poisson variate such that $3P(x=4) = \frac{1}{2} P(x=2) + P(x=0)$

find (i) The mean (ii) $P(x \leq 2)$

sol: Given

$3P(x=4) = \frac{1}{2} P(x=2) + P(x=0)$

$$\frac{3e^{-\lambda}\lambda^4}{4!} = \frac{1}{2}\frac{e^{-\lambda}\lambda^2}{2!} + \frac{e^{-\lambda}\lambda^0}{0!}$$

$$\cancel{e^{-\lambda}}\left[\frac{\cancel{3}\lambda^4}{2\cancel{4}8}\right] = \cancel{e^{-\lambda}}\left[\frac{\lambda^2}{4} + 1\right]$$

$$\Rightarrow \frac{\lambda^2}{8} = \frac{\lambda^2}{4} + 1$$

$$= \frac{\lambda^4}{8} - \frac{\lambda^2}{4} - 1 = 0$$

$$= \frac{\lambda^4 - 2\lambda^2 - 8}{8} = 0$$

$$\Rightarrow \lambda^4 - 2\lambda^2 - 8 = 0$$

$$\lambda^4 - 4\lambda^2 + 2\lambda^2 - 8 = 0$$

$$\lambda^2(\lambda^2 - 4) + 2(\lambda^2 - 4) = 0$$

$$(\lambda^2 + 2)(\lambda^2 - 4) = 0$$

$$\lambda = \pm 2$$

$$\lambda = 2 \quad (\because \lambda \geq 0)$$

(i) mean $= \lambda = 2$

(ii) $P(x \leq 2) = P(x=0) + P(x=1) + P(x=2)$

$$= \frac{e^{-2} 2^0}{0!} + \frac{e^{-2} 2^1}{1!} + \frac{e^{-3} 2^2}{2!}$$

$= \cancel{0.4998} = 0.66$

6) If 2% of the light bulbs are defective find

(i) atleast 1 defective $P(x \geq 1) = 1 - P(x > 1)$

(ii) Exactly 7 are defective $P(x = 7)$

(iii) $P(1 < x < 8)$ in a Sample of 100 $x = 2, 3, 4, 5, 6, 7$

sol: Given $P = 2\% = 0.02$

$n = 100$

mean $= \lambda = np = 0.02 \times 100$
$= 2$

Scanned with CamScann

(i) atleast 1 defective

$$P(X \geq 1) = 1 - P(X = 0)$$

$$= 1 - \left[\frac{e^{-2} \, 1}{1}\right]$$

$$= 1 - [e^{-2}]$$

$$= 0.864764$$

(ii) exactly 7 are defective

$$P(x = 7) = \frac{e^{-2} \cdot 7}{7!}$$

0.0009

$$= 0.00001$$

(iii) $P(1 < x < 8)$

$$= P(x=2) + P(x=3) + P(x=4) + P(x=5) + P(x=6) + P(x=7)$$

$$= \frac{e^{-2} \cdot 2}{2!} + \frac{e^{-2} \cdot 3}{3!} + \frac{e^{-2} \cdot 4}{4!} + \frac{e^{-2} \cdot 5}{5!} + \frac{e^{-2} \cdot 6}{6!} + \frac{e^{-2} \cdot 7}{7!}$$

$$= 0.593$$

Scanned with CamScann

Fit a ~~Binomial~~ Poisson distribution:

1) Fit a ~~Binomial~~ poisson distribution for the following data and calculate the expected frequencies.

x	0	1	2	3	4
f	109	65	22	3	1

sol: $n=4$, $N=\Sigma f=200$

$$\lambda = \text{mean} = \frac{\Sigma fx}{\Sigma f} = \frac{0\times109+1\times65\times2\times22+3\times3+4\times1}{200}$$

$$= \frac{65+44+9+4}{200}$$

$$=0.61$$

$$\therefore \lambda = 0.61$$

Fit a poisson distribution

x	Observed frequency f	$P(x)=\frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-0.61}(0.61)^x}{x!}$	Theoretical frequency $F(x)=N.P(x) = 200.P(x)$
0	109	$P(0)=\frac{e^{-0.61}(0.61)^0}{0!}=0.543$	$0.543\times200 = 109$
1	65	$P(1)=\frac{e^{-0.61}(0.61)^1}{1!}=0.331$	$66.2\simeq66$
2	22	$P(2)=\frac{e^{-0.61}(0.61)^2}{2!}=0.1011$	$20.2\simeq20$
3	3	$P(3)=\frac{e^{-0.61}(0.61)^3}{3!}=0.0205$	$4.1\simeq4$
4	1	$P(4)=\frac{e^{-0.61}(0.61)^4}{4!}=0.0031$	$0.62\simeq1/2$

2) fit a poisson distribution to the following

x	0	1	2	3	4	5
f	142	156	69	27	5	1

sol: $n = 5$, $N = \Sigma f = 400$

$$\lambda = \text{mean} = \frac{\Sigma fx}{\Sigma f} = \frac{0\times142+1\times156+2\times69+3\times27\times 4\times5+5\times1}{400}$$

$$= \frac{400}{400}$$

$$= 1$$

$$\therefore \lambda = 1$$

Fit a poisson distribution

x	Observed frequency f	$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$ $P(x) = \frac{e^{-1}1^{x}}{x!}$	Theortical frequency $F(x) = N \cdot P(x)$ $= 400 \cdot P(x)$
0	142		147
1	156		147
2	69		74
3	27		25
4	5		6
5	1		1

Geometric Distribution.

A Random Variable X is said to have a geometric distribution if it assumes only non-negative values and its probability mass function is said to be

$$P(X=x) = \begin{cases} q^x p & x = 0,1,2,\ldots \quad 0 \le P \le 1,\ q = 1-P \\ 0 & \text{Otherwise} \end{cases}$$

Mean of Geometric Distribution

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

$$= \sum_{x=0}^{\infty} x\, p(x)$$

$$= \sum_{x=0}^{\infty} x \cdot q^x \cdot p$$

$$= \sum_{x=0}^{\infty} x \cdot q^{x-1}\, q \cdot p$$

$$= pq \sum_{x=1}^{\infty} x\, q^{x-1}$$

$$= pq\left[1 + 2q + 3q^2 + 4q^3 + \cdots\cdots\right]$$

$$= pq\,(1-q)^{-2}$$

$$= \cancel{p}q \cdot \frac{1}{p\cancel{2}} \qquad \left[\because 1 + 2x + 3x^2 + \cdots\cdots = (1-x)^{-2}\right]$$

$$= \frac{q}{p}$$

$$\mu = \frac{q}{p}$$

Variance of Geometric Distribution.

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

Consider $E(X^2) = \sum_{x=0}^{\infty} x^2 p(x)$

$$= \sum_{x=0}^{\infty} [x(x-1)+x] p(x)$$

$$= \sum_{x=0}^{\infty} x(x-1) q^x p + \sum_{x=0}^{\infty} x \cdot q^x \cdot p$$

$$= \sum_{x=0}^{\infty} x(x-1) q^{x-2} q^2 p + \frac{q}{p}$$

$$= 2q^2 p \sum_{x=2}^{\infty} \frac{x(x-1)}{2} q^{x-2} + q/p$$

$$= 2q^2 p [1 + 3q + 6q^2 + \dots\dots] + \frac{q}{p}$$

$$= 2q^2 p (1-q)^{-3} + q/p$$

$$= 2q^2 \cancel{p} \cdot \frac{1}{p^{\cancel{3}2}} + \frac{q}{p}$$

$$E(X^2) = 2\frac{q^2}{p^2} + \frac{q}{p}$$

$$V(X) = \frac{2q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2}$$

$$= \frac{q^2}{p^2} + \frac{q}{p}$$

$$V(X) = \frac{q^2 + pq}{p^2}$$

$$= \frac{q(q+p)}{p^2} \quad (\because p+q=1)$$

$$V(X) = q/p^2$$

Scanned with CamScann

Moment generating function of Geometric distribution

$$M_X(t) = E(e^{tx})$$

$$= \sum_{x=0}^{\infty} e^{tx}\, p(x)$$

$$= \sum_{x=0}^{\infty} e^{tx} . q^{x} . p$$

$$= \sum_{x=0}^{\infty} p\,(qe^{t})^{x}$$

$$= p\left[1 + qe^{t} + (qe^{t})^{2} + \dots\right]$$

$$= p\,[1-qe^{t}]^{-1}$$

18/1/2017 Continous probability distribution:

when a Random Variable X takes every value in an interval to gives rise to Continous distribution of X. The distribution defined by Variates like temperature, heights and weights are continous distributions.

probability density function:

for Continous Variable the probability distribution is called probability density function because it is defined for every point in the range and not only f Certain Values. and it is denoted by $f(x)$

Scanned with CamScann

properties:

1) $f(x) \geq 0 \; ; \; -\infty < x < \infty$

2) $\int_{-\infty}^{\infty} f(x)dx = 1$ → Substitute given limits.

3) $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$

Probability distribution function:

Distribution function of a Continuous Random Variable X is denoted by $F(x)$ and is defined as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

properties:

1) $0 \leq F(x) \leq 1, \quad -\infty < x < \infty$

2) $F'(x) = f(x)$

3) $F(-\infty) = 0$

4) $F(\infty) = 1$

Scanned with CamScann

Measures of Central tendency for continous probability distribution:

(i) Mean:

Mean of the distribution is given by

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

If x is defined by A to B then

$$\mu = E(x) = \int_{a}^{b} x\, f(x)\, dx$$

In General

Mean or expection of any function $\phi(x)$ is given by

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

(ii) Median:

is the point which ÷ the entire distribution into 2 equal parts:

In case of Continous distribution median is the point which divides the total area into 2 equal parts.

∴ Thus if X is defined from A to B and M is the median then

$$\int_{a}^{M} f(x)\, dx = \int_{a}^{b} f(x)\, dx = \frac{1}{2}$$

Solving M we get the median.

(iii) Mode:

is the Value of x for which $f(x)$ is maximum.

mode is thus given by

$$f'(x) = 0 \quad \text{and} \quad f''(x) < 0 \quad \text{for } a < x < b$$

(iv) Variance:

$$\sigma^2 = V(x) = \int_{+\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

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

$$= E(x^2) - (E(x))^2$$

(v) Mean Deviation:

Mean Deviation about the mean μ is given by

$$\int_{-\infty}^{\infty} |x-\mu| f(x)\,dx$$

problems:

1) If a probability density of a Random variable is given by $f(x) = \begin{cases} K(1-x^2) & \text{for } 0<x<1 \\ 0 & \text{otherwise} \end{cases}$

find the value of K and the probabilities that a random variable having this probability density will take on value (i) Between 0.1 and 0.2

(ii) >0.5

sol: Given

$$f(x) = \begin{cases} K(1-x^2) & \text{for } 0<x<1 \\ 0 & \text{otherwise} \end{cases}$$

we have $\int_{-\infty}^{\infty} f(x)dx = 1$

i.e $\int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{\infty} f(x)dx = 1$

$$0 + \int_{0}^{1} K(1-x^2)dx + 0 = 1$$

$$K\left[x - \frac{x^3}{3}\right]_0^1 = 1$$

$$K\left[\left(1-\tfrac{1}{3}\right) - (0-0)\right] = 1$$

$$K - 2/3 = 1$$

$$K = 3/2$$

$$\therefore\ f(x) = \begin{cases} 3/2\,(1-x^2) & \text{for } 0<x<1 \\ 0 & \text{otherwise} \end{cases}$$

(i) The probability that the variate will take on a value between 0.1 and 0.2 is

$$P(0.1<x<0.2) = \int_{0.1}^{0.2} f(x)\,dx$$

$$= \int_{0.1}^{0.2} \frac{3}{2}(1-x^2)\,dx$$

$$= \frac{3}{2}\left[x - \frac{x^3}{3}\right]_{0.1}^{0.2}$$

$$= \frac{3}{2}\left[\left[0.2 - \frac{(0.2)^3}{3}\right] - \left[0.1 - \frac{(0.1)^3}{3}\right]\right]$$

$$= 0.1465$$

(ii) $$P(x>0.5) = \int_{0.5}^{\infty} f(x)\,dx$$

$$= \int_{0.5}^{1} f(x)\,dx + \int_{1}^{\infty} f(x)\,dx$$

$$= \int_{0.5}^{1} \frac{3}{2}(1-x^2)\,dx + 0$$

$$= \frac{3}{2}\left[\left(x - \frac{x^3}{3}\right)\right]_{0.5}^{1}$$

$$= 0.312$$

2) The probability density $f(x)$ of a continuous Random variable is given by $f(x) = ce^{-|x|}$, $-\infty < x < \infty$. Show that $c = \frac{1}{2}$ and find the mean and variance of the distributions. Also find the probability that the variate lies between 0 and 4.

Sol :-

Given $f(x) = ce^{-|x|}$

[even functions]

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

$$\Rightarrow \int_{-\infty}^{\infty} ce^{-|x|}\,dx = 1$$

$$\Rightarrow 2c\int_{0}^{\infty} e^{-|x|}\,dx = 1 \quad \left[\text{Since } e^{-|x|} \text{ is even}\right]$$

$$\Rightarrow 2c\int_{0}^{\infty} e^{-x}\,dx = 1 \quad \left[|x| = x\right]$$

$$\Rightarrow 2c\left[\frac{e^{-x}}{-1}\right]_0^{\infty} = 1$$

$$\Rightarrow -2c\left[0-1\right] = 1$$

$$\Rightarrow 2c = 1$$

$$\Rightarrow c = 1/2$$

(ii) mean $= \mu = E(x) = \int_{-\infty}^{\infty} x f(x)dx$

$$= \int_{-\infty}^{\infty} x\, e^{-|x|}\, dx \quad \left[\int_{-a}^{a} f(x)dx = 0 \text{ if } f(x) \text{ is odd} \right.$$

$$\left. = 2\int_{-a}^{a} f(x)dx \text{ is } f(x) \text{ is even} \right]$$

$$= 0 \left[\text{Integrand is odd}\right]$$

(iii) Variance

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

$$= \int_{-\infty}^{\infty} (x-0)^2 \cdot \frac{1}{2} e^{-|x|} dx$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} x^2 e^{-|x|} dx$$

$$= 2 \cdot \frac{1}{2}\int_{0}^{\infty} x^2 e^{-|x|} dx, \text{ Since integrand is even}$$

opposite signs

[ILATE By parts]

$$= \int_{0}^{\infty} x^2 e^{-x} dx = \left(x^2 \frac{e^{-x}}{-1} - 2x\frac{e^{-x}}{1} + 2\frac{e^{-x}}{-1} \right)_0^{\infty}$$

diff

$$= [0 - (-2)] = 2$$

23/1/17

(iii) The probability between 0 and 4 is $= P(0 \le x \le 4)$

$$= \frac{1}{2}\int_0^4 e^{-|x|}dx = \frac{1}{2}\int_0^4 e^{-x}dx$$

$[\because \text{ in } 0 < x < 4, |x| = x]$

$$= \frac{-1}{2}(e^{-x})_0^4$$

$$= -\frac{1}{2}(e^{-4}-1)$$

$$= \frac{1}{2}(1-e^{-4})$$

$= 0.4908$ (Ans)

3) A Continous Random Variable has Probability density function

★★★★★ V. Imp

★

$$f(x) = \begin{cases} kxe^{-\lambda x} & \text{for } x \geq 0, \lambda > 0 \\ 0 & \text{otherwise} \end{cases}$$

λ is Paramete

Determine

(ii) Mean (i) K (iii) Variance

sol: we know $\int_{-\infty}^{\infty} f(x)dx = 1$ [$\because$ Since the total Probabi is unity]

i.e $\int_{-\infty}^{0} 0 \cdot dx + \int_0^\infty kxe^{-\lambda x}dx = 1$ i.e $k\int_0^\infty xe^{-\lambda x}dx = 1$

i.e $k\left[x\left(\frac{e^{-\lambda x}}{-\lambda}\right) - 1\left(\frac{e^{-\lambda x}}{\lambda^2}\right)\right]_0^\infty = 1$.

i.e $K\left[(0-0)-\left(0-1/\lambda^2\right)\right]=1$ or $K=\lambda^2$

now $f(x)$ becomes $f(x)=\begin{cases}\lambda^2 xe^{-\lambda x} & \text{for } x\geq 0, \lambda>0\\ 0, & \text{otherwise}\end{cases}$

(ii) mean of the distribution, $\mu=\int_{-\infty}^{\infty} x\, f(x)\,dx$

$$\text{i.e., } \mu=\int_{-\infty}^{0} 0\cdot dx+\int_{0}^{\infty} x\cdot\lambda^2 xe^{-\lambda x}\,dx=\lambda^2\int_{0}^{\infty} x^2 e^{-\lambda x}\,dx$$

$$=\lambda^2\left[x^2\left(\frac{e^{-\lambda x}}{-\lambda}\right)-2x\left(\frac{e^{-\lambda x}}{\lambda^2}\right)+2\left(\frac{e^{-\lambda x}}{-\lambda^3}\right)\right]_0^{\infty}$$

$$=\lambda^2\left[(0-0+0)-\left(0-0-2/\lambda^3\right)\right]=\frac{2}{\lambda}$$

(iii) Variance of the distribution,

$$\sigma^2=\int_{0}^{\infty} x^2 f(x)\,dx-\mu^2$$

$$\text{i.e., } \sigma^2=\int_{0}^{\infty} x^2 f(x)\,dx-\left(2/\lambda\right)^2=\lambda^2\int_{0}^{\infty} x^3 e^{-\lambda x}\,dx-\frac{4}{\lambda^2}$$

$$=\lambda^2\left[x^3\left(\frac{e^{-\lambda x}}{-\lambda x}\right)-3x^2\left(\frac{e^{-\lambda x}}{\lambda^2}\right)+6x\left(\frac{e^{-\lambda x}}{-\lambda^3}\right)-6\left(\frac{e^{-\lambda x}}{\lambda^4}\right)\right]_0^{\infty}-\frac{4}{\lambda^2}$$

$$=\lambda^2\left[(0-0+0-0)-\left(0-0+0-6/\lambda^4\right)\right]-\frac{4}{\lambda^2}$$

$$=\frac{6}{\lambda^2}-\frac{4}{\lambda^2}$$

$$=\frac{2}{\lambda^2}$$

23

3) A Continuous Random Variable X has the distribution function $F(X)= \begin{cases} 0 & \text{if } x \le 1 \\ k(x-1)^4 & \text{if } 1 < x \le 3 \\ 1 & \text{if } x > 3 \end{cases}$

Determine (i) $f(x)$ (ii) K (iii) mean

$F(X) \to$ PDF
$f(x) \to$ P Den

sol:- (i) we know that

$$f(x) = \frac{d}{dx} F(x)$$

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

(ii) we know that

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

$$\Rightarrow \int_{-\infty}^{1} f(x)\,dx + \int_{1}^{3} f(x)\,dx + \int_{3}^{\infty} f(x)\,dx$$

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

$$\Rightarrow \cancel{4}k\left[\frac{(x-1)^4}{\cancel{4}}\right]_1^3 = 1$$

$$\Rightarrow k(16) = 1$$

$$k = 1/16$$

$k\left[(3-1)^4 - (\cancel{3}-1)^4\right] = 1$

$k(16) = 1$

$k = 1/16$

hence

$$f(x) = \begin{cases} 0 & \text{if } x \leq 1 \\ \frac{1}{4}(x-1)^3 & \text{if } 1 < x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

(iii) mean $= \mu = \int_{-\infty}^{\infty} x f(x)\, dx$

$$= \int_{-\infty}^{1} x f(x)\, dx + \int_{1}^{3} x f(x)\, dx + \int_{3}^{\infty} x f(x)\, dx$$

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

Let $x - 1 = t$

$dx = dt$

upper limit $\Rightarrow x = 3$

$\Rightarrow t = 2$

lower limit $\Rightarrow x = 1$

$\Rightarrow t = 0$

$$\Rightarrow \frac{1}{4}\int_{0}^{2} (t+1)t^3\, dt$$

$$\Rightarrow \frac{1}{4}\int_{0}^{2} (t^4 + t^3)\, dt$$

$$\Rightarrow \frac{1}{4}\left[\frac{t^5}{5} + \frac{t^4}{4}\right]_0^2$$

$$\Rightarrow \frac{1}{4}\left(\frac{32}{5}+\frac{16}{4}\right)$$

$$\Rightarrow \frac{1}{4}\left[\frac{128+80}{20}\right]$$

$$\Rightarrow \frac{1}{4}\left[\frac{208}{20}\right]$$

$$\Rightarrow \frac{52}{20}$$

$$\Rightarrow \frac{10.4}{4}$$

$$\Rightarrow 2.6$$

$\therefore$ mean $= 2.6$

Theorems:

1) If X is a ~~Continous~~ Random Variable and Y is equal to ~~$a(x)+b$~~ $ax+b$ and prove that $E(y) = aE^{x} + b$ and

*** ** v.Imp. Variance of Y is equal to a^2 (Variance of x) i.e $a^2v(x)$ and a, b are constants.

$$V(aX) = a^2 v(x)$$

proof:-

Given $Y = ax+b$ —— ①

$$E(Y) = E(ax+b)$$

$$= \int_{-\infty}^{\infty} (ax+b) f(x) dx$$

$$= \int_{-\infty}^{\infty} ax f(x) dx + \int_{-\infty}^{\infty} b f(x) dx$$

$$= a \int_{-\infty}^{\infty} x f(x) dx + b \int_{-\infty}^{\infty} f(x) dx$$

$E(Y) = aE(x) + b$ —— ②

① − ② gives

$$Y - E(Y) = (ax+b) - (aE(x)+b)$$

$$Y - E(Y) = ax + \cancel{b} - aE(x) - \cancel{b}$$

$$Y - E(Y) = a[x - E(x)]$$

Squaring on both sides

$$[Y - E(Y)]^2 = a^2 (x - E(x)]^2$$

Taking expectation on both sides

$$E[Y - E(Y)]^2 = a^2 E(x - E(x))^2$$

$$V(Y) = a^2 v(x)$$

2) If X is a Continuous Random Variable and K is a constant then prove that

(i) $\text{Variance}(X+K) = V(X)$

(ii) $V(KX) = K^2 \cdot V(X)$

sol:-

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

(i) $$V(X+K) = \int_{-\infty}^{\infty} (x+K)^2 f(x)\,dx - \int_{-\infty}^{\infty} (x+K) f(x)\,dx$$

$$\Rightarrow \int_{-\infty}^{\infty} (x^2 + 2Kx + K^2) f(x)\,dx - \left[\int_{-\infty}^{\infty} x f(x)\,dx + K\int_{-\infty}^{\infty} f(x)\,dx\right]$$

$$\Rightarrow \int_{-\infty}^{\infty} x^2 f(x)\,dx + 2K\int_{-\infty}^{\infty} x f(x)\,dx + K^2\int_{-\infty}^{\infty} f(x)\,dx - \left[\int_{-\infty}^{\infty} x f(x)\,dx - K\right]^2$$

$$\Rightarrow E(X^2) + 2K\,E(X) + K^2 - \left[E(X)+K\right]^2$$

$$V(X+K) = E(X^2) + 2KE(x) + \cancel{K^2} - [E(X)]^2 - 2K\,E(X) - \cancel{K^2}$$

$$= E(X^2) - (E(X))^2$$

$$V(X+K) = V(X)$$

(ii) $Var(kx) = \int_{-\infty}^{\infty} k^2x^2 f(x)dx - \left(\int_{-\infty}^{\infty} kx f(x)dx\right)^2$

$= k^2\int_{-\infty}^{\infty} x^2 f(x)\,dx - k^2\left[\int_{-\infty}^{\infty} x f(x)dx\right]^2$

$= k^2 E(x^2) - k^2(E(x))^2$

$= k^2[E(x^2) - (E(x))^2)$

$Var(kx) = k^2 v(x)$

4) for the Continous Probability function $f(x) = k\cdot x^2 e^{-x}$ when $x \geq 0$ find (i) k (ii) mean (iii) Variance

Sol:- (i) we have $\int_{-\infty}^{\infty} f(x)dx = 1$ $\therefore \int_0^{\infty} kx^2e^{-x}dx = 1 \; (\because x \geq 0)$

i.e., $k\left[x^2(-e^{-x}) - 2x(e^{-x}) + 2(-e^{-x})\right]_0^{\infty} = 1$

$k\left[-e^{-x}(x^2+2x+2)\right]_0^{\infty} = 1$

$k[0+2] = 1$

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

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

$= \int_0^{\infty} kx^3e^{-x}dx$

$= k\left[x^3(-e^{-x}) - 3x^2(e^{-x}) + 6x(-e^{-x}) - 6e^{-x}\right]_0^{\infty}$

$= k\left[-e^{-x}[x^3+3x^2+6x+6\right]_0^{\infty}$

$= k[0+6] = 6k$

$\therefore \mu = 6\left(\frac{1}{2}\right) = 3 \quad \left(\because k = \frac{1}{2}\right)$

24

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

$= \int_{0}^{\infty} x^2 \cdot kx^2 e^{-x}\,dx - (3)^2$

$= k\int_{0}^{\infty} x^4 e^{-x}\,dx - 9$

$= k\left[x^4(-e^{-x}) - 4x^3(e^{-x}) + 12x^2(-e^{-x}) - 24x\,(e^{-x}) + 24\,e^{-x}\right]_0^{\infty} - 9$

$= \frac{1}{2}\left[-e^{-x}\left(x^4 + 4x^3 + 12x^2 + 24x + 24\right)\right]_0^{\infty} - 9$

$= \frac{1}{2}\left[0 + 24\right]$

$= 12 - 9$

$= 3$

24-1-17 Normal Distribution:

A Random Variable X is said to have a normal distribution if its density function or probability distribution is given by

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}$$

$-\infty < x < \infty$

$-\infty < \mu < \infty \quad \sigma > 0$

where

Standard deviation

(mean) μ, σ are 2 parameters of the normal distribution.

Constants of Normal distribution:

(i) Mean

Consider the normal distribution with b, σ as the Parameters then $f(x; b; \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-b}{\sigma}\right)^2}$

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

$$= \int_{-\infty}^{\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-b}{\sigma}\right)^2} dx$$

Let $z = \frac{x-b}{\sigma}$

$x = b + \sigma z$

$dx = \sigma\, dz$

$$\mu = \int_{-\infty}^{\infty} (b+\sigma z)\frac{1}{\cancel{\sigma}\sqrt{2\pi}} e^{-z^2/2} \cancel{\sigma}\, dz$$

$$= \frac{1}{\sqrt{2\pi}}\left[\int_{-\infty}^{\infty} b e^{-z^2/2} dz + \int_{-\infty}^{\infty} \sigma z e^{-z^2/2} dz\right]$$

$$= \frac{1}{\sqrt{2\pi}}\left[b\int_{-\infty}^{\infty} e^{-z^2/2} dz + \sigma\int_{-\infty}^{\infty} z e^{-z^2/2} dz\right]$$

$$= \frac{1}{\sqrt{2\pi}}\left[2b\int_{0}^{\infty} e^{-z^2/2} dz + 0\right] \quad \left[\because e^{-z^2/2} \text{ is even. } z e^{-z^2/2} \text{ is odd}\right]$$

$$\mu = \frac{2b}{\sqrt{2\pi}}\int_{0}^{\infty} e^{-z^2/2} dz$$

$$= \frac{\cancel{2}b}{\sqrt{\cancel{2\pi}}}\sqrt{\frac{\cancel{\pi}}{\cancel{2}}} \quad \left[\because \int_{0}^{\infty} e^{-x^2/2} dx = \sqrt{\pi/2}\right]$$

$$\mu = b$$

(ii) Variance of Normal Distributions:

$$V(x) = E(x-b)^2$$

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

$$= \int_{-\infty}^{\infty} (x-b)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-b}{\sigma}\right)^2} dx$$

Let $Z = \frac{x-b}{\sigma}$

$x-b = \sigma Z$

$dx = \sigma\,dZ$

$$V(x) = \int_{-\infty}^{\infty} (\sigma Z)^2 \frac{1}{\cancel{\sigma}\sqrt{2\pi}} e^{-Z^2/2} \cancel{\sigma}\,dZ$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 Z^2 e^{-Z^2/2} dZ$$

$$= \frac{\sigma^2}{\sqrt{2\pi}}\, 2 \int_{0}^{\infty} Z^2 e^{-Z^2/2} dZ$$ [Since Integrand is even]

put $Z^2/2 = t$

$\cancel{2} Z\,dZ = dt$

$\cancel{2} dZ = \frac{dt}{Z} = \frac{dt}{\sqrt{2t}}$

$$V(x) = \frac{2\sigma^2}{\sqrt{2\pi}} \int_0^\infty 2t\, e^{-t} \frac{dt}{\sqrt{2t}}$$

$$= \frac{2\sigma^2}{\sqrt{2\pi}} \int_0^\infty \sqrt{2t}\, e^{-t} dt$$

$\boxed{\Gamma = \gamma}$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^\infty t^{1/2} e^{-t} dt$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^\infty e^{-t}\, t^{3/2-1} dt \qquad \left[\int_0^\infty e^{-x} x^{n-1} dx = \gamma_n \right]$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \gamma_{3/2} \qquad \Gamma{n+1} = n\Gamma{n}$$

$$V(x) = \frac{2\sigma^2}{\sqrt{\pi}} \Gamma{1/2+1}$$

$$= \frac{\cancel{2}\sigma^2}{\sqrt{\pi}} \frac{1}{\cancel{2}} \Gamma{1/2} \qquad [\gamma_{n+1} = n\gamma_n]$$

$$= \frac{\sigma^2}{\cancel{\sqrt{\pi}}} \cdot \cancel{\sqrt{\pi}} \qquad [\gamma_{1/2} = \sqrt{\pi}]$$

$$V(x) = \sigma^2$$

Thus the Standard deviation of normal distribution is σ

Mode of Normal Distribution:

Mode is a value of x for which $f(x)$ is maximum, i.e mode is the solution of $f'(x)=0$ and $f''(x)<0$

By definition, we have

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Diff.... wrt 'x' we get

$$f'(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} \cdot -\frac{1}{\cancel{2}} \cdot \cancel{2}\left(\frac{x-\mu}{\sigma}\right) \cdot \frac{1}{\sigma}$$

$$f'(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} \left(-\left(\frac{x-\mu}{\sigma^2}\right)\right)$$

$$f'(x) = -\left(\frac{x-\mu}{\sigma^2}\right) f(x)$$

now $f'(x)=0$

$$-\left(\frac{x-\mu}{\sigma^2}\right) f(x) = 0$$

$$x-\mu = 0$$

$$x = \mu$$

$$f''(x) = -\left[\left(\frac{x-\mu}{\sigma^2}\right) f'(x) + f(x) \cdot \frac{1}{\sigma^2}\right]$$

$$= -\left[\left(\frac{x-\mu}{\sigma^2}\right)\left(-\left(\frac{x-\mu}{\sigma^2}\right) f(x)\right) + f(x)\frac{1}{\sigma^2}\right]$$

$$= \frac{-f(x)}{\sigma^2}\left[1 - \frac{(x-\mu)^2}{\sigma^2}\right]$$

At the point $x = \mu$

$$f''(x) = \frac{-f(x)}{\sigma^2}$$

$$\therefore f''(x) < 0$$

Hence

$x = \mu$ is the mode of the distribution.

Median of the normal distribution

If m is the median of the normal distributio

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

i.e

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{2}$$

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx + \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad \text{--- (1)}$$

now

Consider $\displaystyle \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$

Let $Z = \dfrac{x-\mu}{\sigma}$

$$dx = \sigma\, dZ$$

upper limit $x = \mu \Rightarrow z = 0$

lower limit $x \to -\infty \Rightarrow z \to -\infty$

$$\therefore \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\cancel{\sigma}\sqrt{2\pi}} \int_{-\infty}^{0} e^{-z^2/2} \cancel{\sigma} dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-z^2/2} dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-z^2/2} dz \quad [\text{by symmetric}]$$

$$= \frac{1}{\sqrt{2\cancel{\pi}}} \sqrt{\cancel{\pi}/2}$$

$$= \frac{1}{2} \quad \text{——— (2)}$$

from (1) and (2), we have

$$\frac{1}{2} + \frac{1}{\sigma\sqrt{2\pi}} \int_{dt}^{M} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1/2$$

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} dx = 0$$

$$\therefore \mu = M \quad \left[\text{If } \int_a^b f(x)\,dx = 0 \text{ then } a = b \text{ where } f(x) > 0\right]$$

Hence for the normal distribution mean = median = mode.

Q) we notice that for the normal distributions mean, median and mode Coincide.

Prove it

Hence the distribution is Symmetrical.

Mean deviation from the mean for normal distributions:

By definition, mean deviation about the mean

$$= \int_{-\infty}^{\infty} |x-\mu| f(x) dx$$

$$= \int_{-\infty}^{\infty} |x-\mu| \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2 \left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} |x-\mu| e^{-1/2 \left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Let $z = \frac{x-\mu}{\sigma}$

$x-\mu = \sigma z$

$dx = \sigma dz$

$$\frac{1}{\cancel{\sigma}\sqrt{2\pi}} \int_{-\infty}^{\infty} |\sigma z| e^{-z^2/2} \cancel{\sigma} dz$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |z| e^{-z^2/2} dz$$

$$= \frac{2\sigma}{\sqrt{2\pi}} \int_0^\infty z e^{-z^2/2} \, dz \quad \left[\text{Since Integrand is even}\right]$$

Let $z^2/2 = t$

$$\frac{\cancel{2}z}{\cancel{2}} dz = dt$$

$$z\,dz = dt$$

$$\therefore dz = \frac{dt}{z}$$

$$= \frac{2\sigma}{\sqrt{2\pi}} \int_0^\infty \cancel{z} e^{-t} \frac{dt}{\cancel{z}}$$

$$= \frac{2\sigma}{\sqrt{2\pi}} \left[\frac{e^{-t}}{-1}\right]_0^\infty$$

$$= \sqrt{\frac{2}{\pi}}\, \sigma \left[0+1\right]$$

$$= \sigma\sqrt{\frac{2}{\pi}}$$

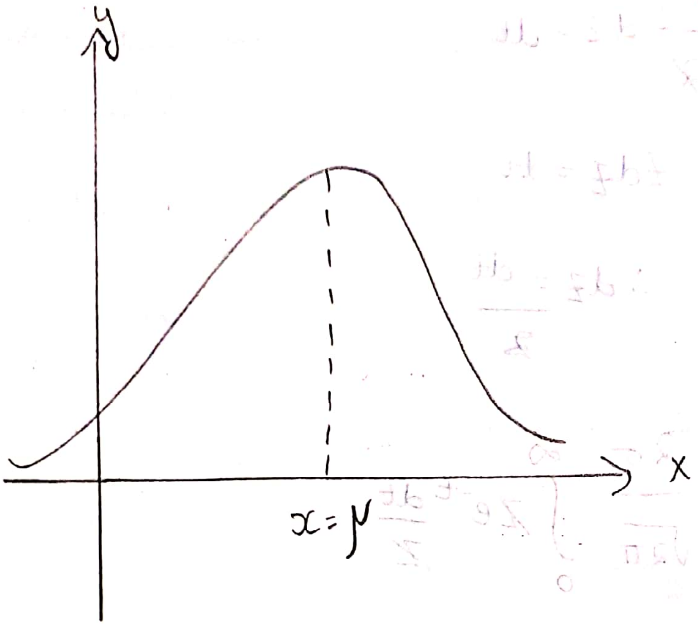
$$\boxed{\sqrt{\frac{2}{\pi}} = \frac{4}{5}}$$

$$= \frac{4}{5}\sigma$$

Hence the Mean deviation about the mean for normal distribution is $\frac{4}{5}\sigma$

**** properties of normal distributions:
v Imp:-

(1) The Graph of the normal distribution $y=f(x)$ in the xy-plane is known as the normal curve.



(2) The curve is a bell-Shaped curve and symmetrical with respect to mean. i.e about the line $x=\mu$ and the two tails on the right and the left sides. of the mean μ extends to infinity. The top of the bell is directly above the mean

(3) Area under the normal curve represents the total population.

(4) Mean, Mode, Median of the distribution Coincide at $x=\mu$ as the distribution is Symmetrical. So Normal curve is unimodal.

(5) X-axis is an asymptote to the curve.

(6) linear Combination of independent normal Variates is also a normal Variate.

Standard Normal distributions.

Normal Distributions with mean $\mu = 0$ and Standard deviation $\sigma = 1$ is known as Standard Normal Distributions

note:-

~~$x = \mu / \sigma$~~ $Z = \frac{x-\mu}{\sigma}$

is called Standard Normal Variate.

25/1/16 1) If X is a normal Variate with mean $\mu = 30$ and Standard deviation 5. find the probabilities that

(i) $26 \le x \le 40$ (ii) $x \ge 45$

sol:- $\mu = 30, \sigma = 5$

Standard deviation $\sigma = 5$

Standard normal Variate $Z = \frac{X-\mu}{\sigma}$

(i) where $X = 26$,

$Z = \frac{26-30}{5} = \frac{-4}{5} = -0.8 = Z_1$

[- or + both are Same]

when $X = 40$,

$Z = \frac{40-30}{5} = \frac{10}{5} = 2 = Z_2$

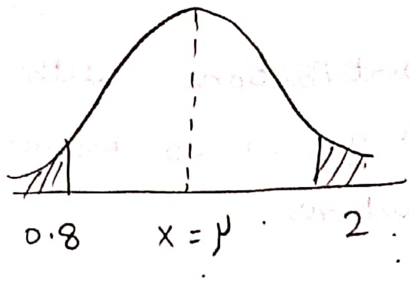
$P(26 \le x \le 40) = P(Z_1 \le Z < Z_2)$

$= P(-0.8 \le Z \le 2)$

$= A(2) + A(0.8)$

A( ) ↓ Area

Scanned with CamScanner



$$P(26 \le x \le 40) = 0.4772 + 0.2881$$

$$= 0.7653$$

(ii) when $x = 45$

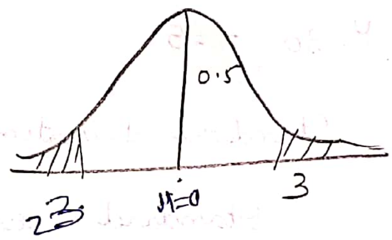
$$Z = \frac{x-\mu}{\sigma} = \frac{45-30}{5} = \frac{15}{5} = 3$$

$$P(x \ge 4.5) = P(Z \ge 3)$$

$$= 0.5 - A(3)$$

$$= 0.5 - 0.4987$$

$$= 0.0013$$



2) If the masses of 300 Students are normally distributed with mean 68 kgs and σ 3 kgs. How many Students have masses?

(i) ≥ 72 kgs

(ii) ≤ 64 kgs

(iii) Between 65 and 75 kgs inclusive

sol:- let μ be the mean, σ is the Standard deviation of the distribution, then

$$\mu = 68 \text{ kgs} \quad \text{and} \quad \sigma = 3 \text{ kgs}$$

Let the Variable x denotes the masses of students

when $x = 72$

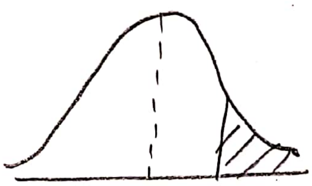
$$Z = \frac{x-\mu}{\sigma} = \frac{72-68}{3} = \frac{4}{3} = 1.33$$

$$P(x > 72) = P(z > 1.33)$$

$$= 0.5 - A(1.33)$$

$$= 0.5 - 0.4082$$

$$= 0.0918$$



1.33

(i) Number of students more than 72 kgs

$$= 300 \times 0.0918$$

$$= 28 \text{ (approximately)}$$

(ii) where $x = 64$

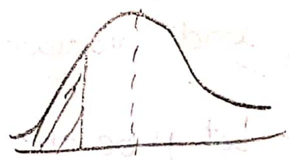
$$Z = \frac{x-\mu}{\sigma} = \frac{64-68}{3} = \frac{-4}{3} = -1.33$$

$$P(x \leq 64) = P(z \leq -1.33)$$

$$= 0.55 - A(1.33)$$

$$= 0.55 - 0.4082$$

$$= +0.0918$$



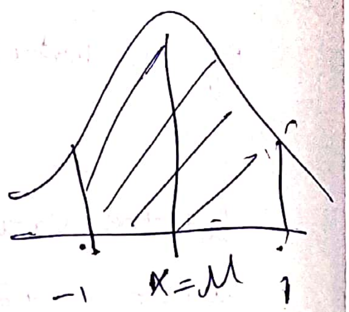
-1.33

no. of students masses less than or equal to 64 kg

$$= 300 \times 0.0918$$

$$= 28$$

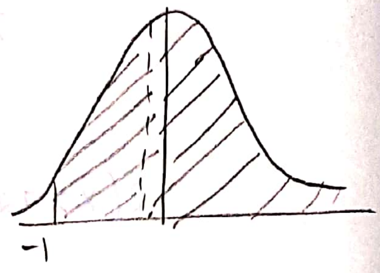
(iii) when $x = 65$



$$Z = \frac{x-\mu}{\sigma} = \frac{65-68}{3} = \frac{-3}{3} = -1$$

when $x = 71$

$$Z = \frac{x-\mu}{\sigma} = \frac{71-68}{3} = \frac{3}{3} = 1$$



$$P(65 < x < 71) = P(-1 < Z < 1)$$

$$= A(1) + A(1)$$

$$= 2A(1)$$

$$= 2(0.3413)$$

$$= 0.6826$$

$\Rightarrow$ no. of Students whose weight lie between $65 \le x \le 71$

$$300 \times 0.6826$$

$$= 205$$

★★★★★ V.Imp Sum + model

3) In a Normal distribution 7% of the items under 35 and 89% are under 63. Determine the mean and Variance of the distribution.

Sol:- Let μ be the mean (at $Z=0$) and σ be the Standard deviation of the normal curve

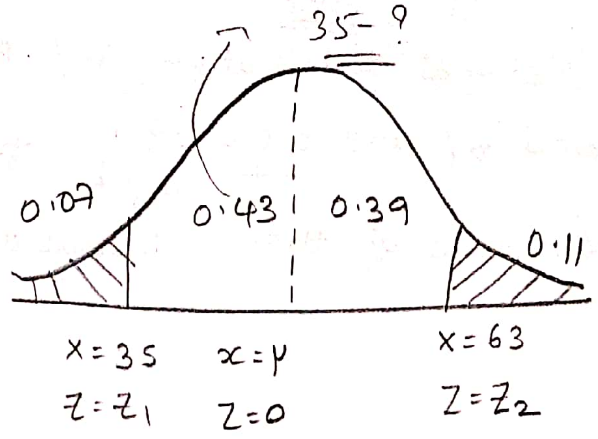
7% of the items under 35 means the area to the left of the ordinate at $x = 35$

Given $P(X<35) = 0.07$

and $P(X<63) = 0.89$

$$P(X>63) = 1 - P(X<63)$$

$$= 1 - 0.89 = 0.11$$



where $x=35$, $Z=\frac{x-\mu}{\sigma}=\frac{35-\mu}{\sigma}=-Z_1$ — (1)

where $x=63$, $Z=\frac{x-\mu}{\sigma}=\frac{63-\mu}{\sigma}=Z_2$ — (2)

from the diagram, we have

$P(0<Z<Z_1)=0.39 \Rightarrow Z_2=1.23$ [from Table Values]

$P(0<Z<Z_2)=0.43 \Rightarrow Z_1=1.48$

from (1) $\frac{35-\mu}{\sigma}=-1.48 \Rightarrow 35-\mu=-1.48\sigma$

$\Rightarrow \mu-1.48\sigma=35$ — (3)

from (2) $\frac{63-\mu}{\sigma}=1.23$

$63-\mu=1.23\sigma$

$\mu+1.23\sigma=63$ — (4)

on solving (3) and (4)

$\mu=50.3$, $\sigma=10.332$

HW

4) In a Normal distribution 31% of the items are under 45 and 8% are over 64 to the right. Find the mean and variance of the distribution?

sol:-

28/1/17

5) For a normally distributed Variate with mean 1 and σ 3. find the probabilities that

(i) $3.43 \le x \le 6.19$

(ii) $-1.43 \le x \le 6.19$

6) Given that the mean height of Students in a class is 158 Cms with σ of 20 cms. find how many Student lies between 150 Cms and 170 Cms. If there are 100 Students in a class?

Solutions:-

(4) Sol: Let X be the Continuous random Variable.

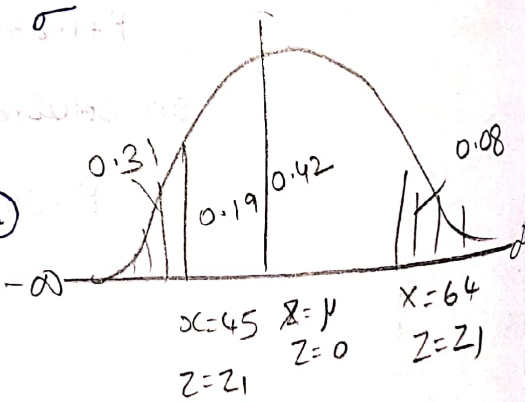
Let μ be the mean and σ the Standard deviation

Given $P(X<45) = 0.31$ and $P(X>64) = 0.08$

Standard variable, $Z = \dfrac{X-\mu}{\sigma}$

when $X = 45$, Let $Z = Z_1$

So that $Z_1 = \dfrac{45-\mu}{\sigma}$ —(1)



$$\therefore \int_{-\infty}^{Z_1} \phi(z)\, dz = 0.31$$

$$= \int_{-\infty}^{0} \phi(z)\, dz - \int_{Z_1}^{0} \phi(z)\, dz = 0.31$$

Scanned with CamScanner

$$\therefore \int_{Z_1}^{0} \phi(Z)\,dZ = \int_{-\infty}^{0} \phi(Z)\,dZ - 0.31 = 0.5 - 0.31 = 0.19$$

Hence $P(0 < Z < Z_1) = 0.19 \Rightarrow Z_1 = -0.5$ (from table) — ②

when $X = 64$, $Z = \dfrac{64-\mu}{\sigma} = Z_2$ (say)

$$\therefore \int_{Z_2}^{\infty} \phi(Z)\,dZ = 0.08 \text{ or } \int_{0}^{\infty} \phi(Z)\,dZ - \int_{0}^{Z_2} \phi(Z)\,dZ = 0.08$$

Hence $$\int_{0}^{Z_2} \phi(Z)\,dZ = \int_{0}^{\infty} \phi(Z)\,dZ - 0.08 = 0.5 - 0.08 = 0.42$$

Thus $P(0 < Z < Z_2) = 0.42$

$Z_2 = 1.4$ (from tables) — ④

from ① and ②, we have

$$\frac{45-\mu}{\sigma} = -0.5 \Rightarrow 45 - \mu = -0.5\sigma \quad \text{— ⑤}$$

from ③ and ④ we get

$$\frac{64-\mu}{\sigma} = 1.4 \Rightarrow 64 - \mu = 1.4\sigma \quad \text{— ⑥}$$

⑤ - ⑥ gives $(45-\mu) - (64-\mu) = -0.5\sigma - 1.4\sigma$

$\Rightarrow -19 = -1.9\sigma$

$$\therefore \sigma = \frac{19}{1.9} = 10$$

from (5), $\mu = 45 + 0.5\sigma = 45 + 0.5(10) = 50$

Hence mean $= 50$ and $\sigma = 10$

(5)

Sol:- Given $\mu = 1$ and $\sigma = 3$

(i) when $x = 3.43$, $Z = \frac{x-\mu}{\sigma} = \frac{3.43-1}{3} = \frac{2.43}{3}$

$= 0.81 = Z_1$ (say)

when $x = 6.19$, $Z = \frac{x-\mu}{\sigma} = \frac{6.19-1}{3} = \frac{5.19}{3}$

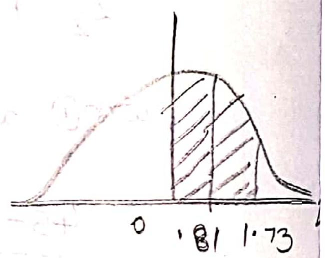
$= 1.73 = Z_2$ (Say)

$\therefore P(3.43 \le x \le 6.19) = P(0.81 \le Z \le 1.73)$

$= A(Z_2) - A(Z_1)$

$= A(1.73) - A(0.81) = 0.458 - 0.291$ (from Tables)

$= 0.1672$



(ii) when $x = -1.43$, $Z = \frac{x-\mu}{\sigma} = \frac{-1.43-1}{3} = -0.81 = Z_1$ (Say)

when $x = 6.19$, $Z = \frac{x-\mu}{\sigma} = \frac{6.19-1}{3} = 1.73 = Z_2$ (Say)

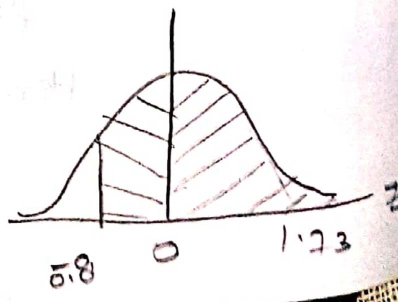
$\therefore P(-1.43 \le x \le 6.19) = P(-0.81 \le Z \le 1.73)$

$= A(1.73) + A(-0.81)$

$= A(1.73) + A(0.81)$ $(\because A(-Z) = A(Z))$

$= 0.4582 + 0.2910$

$= 0.7942$



(6)

sol

⑥ HW

sol' we have

mean, $\mu = 158$ cms and

Standard deviation $\sigma = 20$ cms

$$\therefore Z = \frac{X-\mu}{\sigma}$$

$$= \frac{X-158}{20}$$

when $X = 150$,

$$Z = \frac{150-158}{20}$$

$$= \frac{-2}{5} \Rightarrow -0.4$$

when $X = 170$,

$$Z = \frac{170-158}{20}$$

$$= \frac{3}{5}$$

$$= 0.6$$

$$\therefore P(150 \leq X \leq 170) = P(-0.4 \leq Z \leq 0.6)$$

$$= P(-0.4 \leq Z \leq 0) + P(0 \leq Z \leq 0.6)$$

$$= P(0 \leq Z \leq 0.4) + P(0 \leq Z \leq 0.6),$$

[due to Symmetry]

$= 0.1554 + 0.2257$

$= 0.3811$

number of students whose height lie between 150 cms and 170 cms

$=$ Probability $\times$ total number of students.

$= 0.3811 \times 100$

$= 38$

($\because$ number of students should be integer.

1

SAMPLING DISTRIBUTION

Population:- A group of individuals under study is called population (or) Population is the collection of objects.

A Population may be finite or infinite. If the population contains a finite no. of units then it is called finite population.

Ex:- No. of students in a college.

If the population contains an infinite no. of units called infinite population.

Ex:- Stars in the sky

$x_1, x_2, \ldots\ldots x_N$ are population units.

Sample:- A finite subset of statistical individuals in a population is called a sample (or) subset of a population.

Ex:- Blood ~~sam~~testing

Sample size:- The no. of objects in the sample is called sample size. and it is denoted by n.

NOTE:- Population size is denoted by N.

TYPES OF SAMPLING:-

There are four types of sampling

1. Purposive sampling
2. Random sampling
3. Stratified sampling
4. Systematic sampling

① <u>Purposive Sampling</u>:- If the sample elements are selected with a definite purpose in mind, then the sample selected is called purposive sampling.

② Random sampling (or Probability sampling):- It is the process of drawing a sample from a population in such a way that each member of the population has an equal chance of being included in the sample. The sample obtained by the process of random sampling is called a random sample.

Ex:- Selecting randomly 20 words from a dictionary is a random sample.

If each element of a population may be selected more than once then it is called sampling with replacement whereas if the element cannot be selected more than once, it is called sampling without replacement.

<u>NOTE</u>:- If N is the size of a population and n is the sample size, then

(i) The number of samples with replacement $= N^n$

(ii) The number of samples without " $= {}^{N}C_n$

③ <u>Stratified sampling</u>:- This method is useful when the population is heterogeneous. In this type of sampling, the population is first sub-divided into several parts (or small groups) called strata according to some relevant characteristics so that each stratum is more or less homogeneous. Each stratum is called a sub-population. Then a small sample (called sub-sample) is selected from each stratum at random. All the sub-samples are combined together to form the stratified sample which represents the population properly.

The Process of obtaining and obtaining a stratified sample with a view to estimating the characteristic of the population is known as stratified sampling.

2

④ Systematic sampling:- If the Population size is finite, all the units of the population are arranged in some order, Then from the first k items, one unit is selected at random. This unit and every kth unit of the serially listed population combined together constitute a systematic sample. This type of sampling is known as systematic sampling.

CLASSIFICATION OF SAMPLES:-

Samples are classified in two ways.

1) Large Sample: If the size of the sample $(n) \geq 30$, the sample is said to be large sample.

2) Small sample:- If the size of the sample $(n) < 30$, the sample is said to be small sample (or) exact sample.

PARAMETERS AND STATISTICS:-

Statistical measures computed from the sample observations is known as statistic. Let $x_1, x_2 \dots x_n$ are n sample observations.

mean $\bar{x}$ & standard deviation (s) are known as statistic.

In other words, the mean, median, mode, S.D. variance measures of the population are called parameters and the measures obtained from the sample of the population are called statistics.

The parameter refers to population while statistic refer to sample.

Sample mean:- If $x_1, x_2, \ldots x_n$ represents a random sample of size n, then the sample mean is denoted by the statistic

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

Sample variance:- If $x_1, x_2 \ldots x_n$ represents a random of sample size n, then the sample variance is denoted by the statistic.

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

CENTRAL LIMIT THEOREM:-

If $x_i, (i=1,2,\ldots n)$ be independent random variables such that $E(x_i)=\mu_i$ and $V(x_i)=\sigma_i^2$, then under certain very general conditions, the random variable $S_n = X_1 + X_2 + \ldots + X_n$, is asymptotically normal with mean μ and standard deviation σ

where $\mu = \sum_{i=1}^{n} \mu_i$ and $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$

SAMPLING DISTRIBUTION OF MEAN (σ KNOWN):-

The Probability distribution of $\bar{x}$ is called the sampling distribution of means. The sampling distribution of a statistic depends on the size of the population, the size of the samples, and the method of choosing the samples.

Infinite Population:- Suppose the samples are drawn from an infinite population (or) sampling is done with replacement, then

the mean of the sampling distribution of means,

$$\mu_{\bar{x}} = \frac{\mu+\mu+\mu+\ldots+\mu}{n} = \mu$$

and variance $$\sigma_{\bar{x}}^2 = \frac{\sigma^2+\sigma^2+\ldots+\sigma^2}{n} = \frac{\sigma^2}{n}.$$

8

$\therefore$ S.D of mean, $\sigma_{\bar{x}} = \sigma/\sqrt{n}$

The total no. of sample with replacement $= N^n$

The sampling distribution of $\bar{x}$ will be approximately normal with mean μ and variance $\frac{\sigma^2}{n}$ provided that the sample size is large.

Standard sample mean, $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$.

Finite Population:- Consider a finite population of size N with mean μ and standard deviation σ. Draw all possible samples of size n without replacement, from this population.

The mean of the sampling distribution of mean (for $N > n$) is $\mu_{\bar{x}} = \mu$.

The variance is $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \cdot \left(\frac{N-n}{N-1}\right)$

and S.D. is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

Here, the factor $\left(\frac{N-n}{N-1}\right)$, often called the finite population corrector factor.

The number of samples without replacement $= {}^{N}C_{n}$

① Find the value of the finite population correction factor for $n=10$ and $N=1000$.

Sol: Given N = The size of the finite population = 1000

n = The size of the sample = 10

$\therefore$ Correction factor $= \frac{N-n}{N-1} = \frac{1000-10}{1000-1}$

$= \frac{990}{999}$

$= 0.991$.

② How many different samples of size two can be chosen, from a finite population of size 25.

Sol: We can take ^NC_n samples of size 'n' from the population of size N.

Here $N=25, n=2$

$\therefore$ we can take $^{25}C_2 = 300$ samples of size 2 from finite population of size 25.

③ A population consists of five numbers 2, 3, 6, 8 and 11. Consider all possible samples of size two which can be drawn with replacement from this population. Find

a) The mean of the Population.

b) The standard deviation of the population.

c) The mean of the sampling distribution of means and

d) The standard deviation of the sampling distribution of means (i.e., standard error of means).

Sol: a) Mean of the population is given by

$$\mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$$

b) Variance of the population (σ^2) is given by

$$\sigma^2 = \sum \frac{(x_i-\bar{x})^2}{n}$$

$$= \frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5}$$

$$= 10.8$$

$$\therefore \sigma = \sqrt{10.8}$$

i.e., $\sigma = 3.29$

Scanned with CamScanner

(c) Sampling with replacement (Infinite Population):

The total no. of samples with replacement is

$N^n = 5^2 = 25$ samples of size 2.

Here N = Population size and n = Sample size listing all possibles samples of size 2 from population 2, 3, 6, 8, 11 with replacement we get 25 samples.

$$\begin{Bmatrix} (2,2) & (2,3) & (2,6) & (2,8) & (2,11) \\ (3,2) & (3,3) & (3,6) & (3,8) & (3,11) \\ (6,2) & (6,3) & (6,6) & (6,8) & (6,11) \\ (8,2) & (8,3) & (8,6) & (8,8) & (8,11) \\ (11,2) & (11,3) & (11,6) & (11,8) & (11,11) \end{Bmatrix}$$

Now compute the arithmetic mean for each of these 25 samples. The set of 25 means $\bar{x}$ of these 25 samples, gives rise to the distribution of means of the samples known as sampling distribution of means.

The sample mean are

$$\begin{Bmatrix} 2 & 2.5 & 4 & 5 & 6.5 \\ 2.5 & 3 & 4.5 & 5.5 & 7 \\ 4 & 4.5 & 6 & 7 & 8.5 \\ 5 & 5.5 & 7 & 8 & 9.5 \\ 6.5 & 7 & 8.5 & 9.5 & 11 \end{Bmatrix} \text{———} (I)$$

and the mean of sampling distribution of means is the mean of these 25 means.

$$\mu_{\bar{x}} = \frac{\text{sum of all sample means in (I)}}{25} = \frac{150}{25} = 6$$

Illustrating that $\mu_{\bar{x}} = \mu$.

d) The variance $\sigma_{\bar{x}}^2$ of the sampling distribution of means is obtained by subtracting the mean 6 from each number in (I) and squaring the result, adding all 25 members they obtained, and dividing by 25.

$$\sigma_{\bar{x}}^2 = \frac{(2-6)^2 + \cdots\cdots + (11-6)^2}{25} = \frac{135}{25} = 5.40$$

and thus $\sigma_{\bar{x}} = \sqrt{5.40} = 2.32$

Clearly, for finite population involving sampling with replace- -ment (or infinite population)

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \text{ or } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{3.29}{\sqrt{2}} = 2.32.$$

(2) above problem as without replacement

a & b same

(c) Sampling without replacement (finite population):

The total no. of samples without replacement is $Nc_n = 5c_2 = 10$ samples of size 2.

The 10 samples are

$$\left\{\begin{matrix} (2,3) & (2,6) & (2,8) & (2,11) \\ (3,6) & (3,8) & (3,11) & \\ (6,8) & (6,11) & & \\ (8,11) & & & \end{matrix}\right\}$$

The selection (2,3) is considered same as (3,2)

The corresponding sample means are.

$$\left\{\begin{matrix} 2.5 & 4 & 5 & 6.5 \\ 4.5 & 5.5 & 7 & \\ 7 & 8.5 & & \\ 9.5 & & & \end{matrix}\right\}$$

The mean of the sampling distribution of means is

$$\mu_{\bar{x}} = \frac{(2.5+4+5+6.5+4.5+5.5+7+7+8.5+9.5)}{10} = 6$$

Illustrating that $\mu_{\bar{x}} = \mu$

d) The variance of sampling distribution of means

$$\sigma_{\bar{x}}^2 = \frac{(2.5-6)^2+(4-6)^2+\cdots+(9.5-6)^2}{10}$$

$$= 4.05$$

$$\sigma_{\bar{x}} = 2.01$$

Showing that $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$

$$= \frac{10.8}{2}\left(\frac{5-2}{5-1}\right) = 4.05$$

for sampling without replacement.

② Find the mean and standard deviation of sampling distribution of variances for the population 2, 3, 4, 5 by drawing samples of size two (a) with replacement (b) without replacement.

③ A population consists of six numbers 4, 8, 12, 16, 20, 24. Consider all samples of size two which can be drawn without replacement from this population. Find

a) The Population mean

b) The Population Standard deviation

c) The mean of the sampling distribution of means

d) The S.D. of the sampling distribution of means.

④ Samples of size 2 are taken from the population 1, 2, 3, 4, 5, 6 (i) with replacement & (ii) without replacement.

a, b, c, d. (above)

⑤ If the population is 3, 6, 9, 15, 27

a) list all possible sample of size 3 that can be taken without replacement from the finite population

b) Calculate the mean of each of the sampling distribution of means.

c) Find the S.D. of sampling distribution of means

Sol:

Mean of the population $\mu = \frac{3+6+9+15+27}{5} = \frac{60}{5} = 12$

standard deviation of the population.

$$\sigma = \sqrt{\frac{(3-12)^2+(6-12)^2+\cdots+(27-12)^2}{5}}$$

$$= \sqrt{\frac{360}{5}} = 8.4853$$

a) Sampling without replacement (finite population):

The total number of samples without replacement is

$${}^{N}C_{n} = {}^{5}C_{3} = 10$$

The 10 samples are

(3, 6, 9), (3, 6, 15), (3, 9, 15), (3, 6, 27), (3, 9, 27), (3, 5, 27), (6, 9, 15), (6, 9, 27), (6, 15, 27), (9, 15, 27).

b) Mean of the sampling distribution of mean is

$$\mu_{\bar{x}} = \frac{6+8+9+10+12+13+14+15+16+17}{10} = \frac{120}{10} = 12$$

c) $\sigma_{\bar{x}}^2 = \frac{1}{9}\left[(16-12)^2+(8-12)^2+\cdots+(17-12)^2\right]$

$$= \frac{120}{9}$$

$$= 13.3$$

$\therefore \sigma_{\bar{x}} = \sqrt{13.3} = 3.651.$

6

⑥ The mean height of students in a college is 155 cms and standard deviation is 15. What is the probability that the mean height of 36 students is less than 157 cms.

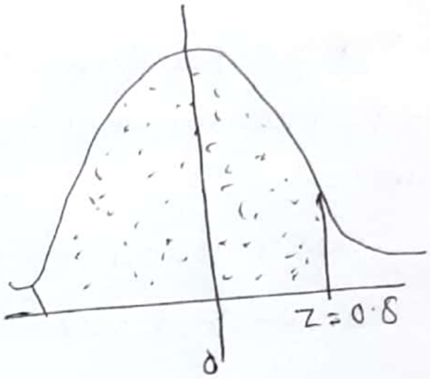
Sol:- μ = mean of the population

= mean height of students of a college = 155cm

σ = S.D. of population = 15cms

n = Sample size = 36

$\bar{x}$ = Mean of sample = 157cms



Now $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

$= \dfrac{157-155}{15/\sqrt{36}} = 0.8$

$\therefore P(\bar{x} \le 157) = P(z<0.8) = 0.5 + P(0 \le z < 0.8)$

$= 0.5 + 0.2881 = 0.7881$

Thus the probability that the mean height of 36 students is less than 157 = 0.7881.

⑦ A random sample of size 100 is taken from an infinite population having the mean $\mu = 76$ and the variance $\sigma^2 = 256$. What is the probability that $\bar{x}$ will be between 75 and 78.

Sol:- n = size of the sample = 100

μ = mean of the population = 76

σ^2 = variance of the population = 256

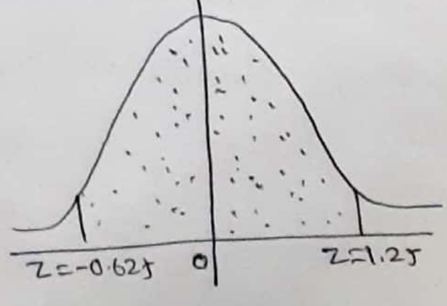
$\therefore \sigma = 16$

Since n is large, the sample mean $\bar{x} \sim N(\mu, \sigma^2/n)$

$\therefore Z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

When $\bar{x}_1 = 75$

$Z_1 = \dfrac{75-76}{16/\sqrt{100}} = -0.625$



and $\bar{x}_2 = 78$, $Z_2 = \frac{\bar{x}_2 - \mu}{\sigma/\sqrt{n}} = \frac{78-76}{16/\sqrt{100}} = 1.25$

$\therefore P(75 \le \bar{x} \le 78) = P(Z_1 \le Z \le Z_2)$

$= P(-0.625 \le Z \le 1.25)$

$= P(-0.625 \le Z \le 0) + P(0 \le Z \le 1.25)$

$= 0.2334 + 0.3944$

$= 0.628$

⑧ A normal population has a mean of 0.1 and S.D. of 2.1. Find the probability that mean of a sample of size 900 will be negative.

Sol:- Given $\mu = 0.1$, $\sigma = 2.1$ and $n = 900$

The standard normal variate

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x} - 0.1}{2/\sqrt{900}}$$

$\therefore \bar{x} = 0.1 + 0.007Z$, where $Z \sim N(0,1)$

$\therefore$ The required probability, that the sample mean is negative is given by

$P(\bar{x} < 0) = P(0.1 + 0.07Z < 0)$

$= P(0.07Z < -0.1)$

$= P\left(Z < \frac{-0.1}{0.07}\right)$

$= P(Z < -1.43)$

$= 0.50 - P(0 < Z < 1.43)$

$= 0.50 - 0.4236$

$= 0.0764.$

⑨ A random sample of size 64 is taken from a normal population with $\mu = 51.4$ and $\sigma = 68$. What is the probability that the mean of the sample will (a) exceed 52.9

(b) fail between 50.5 and 52.3

(c) be less than 50.6.

SAMPLING DISTRIBUTION OF DIFFERNCES AND SUMS:-

Let μ_{S_1} and σ_{S_1} be the mean and S.D. of sampling distribution of statistic S_1 obtained by computing S_1 for all possible samples of size n_1 draw from a population A. Also let μ_{S_2} and σ_{S_2} be the mean and standard deviation of sampling distribution of statistic S_2 obtained by computing S_2 for all possible samples of size n_2 drawn from another different population B.

Now compute the statistic $S_1 - S_2$, the difference of the statistic from all the possible combinations of these samples from the two populations A and B.

Then The mean $\mu_{S_1-S_2}$ and the S.D. $\sigma_{S_1-S_2}$ of the sampling distribution of differences are given by

$$\mu_{S_1-S_2} = \mu_{S_1} - \mu_{S_2} \text{ and } \sigma_{S_1-S_2} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}$$

assuming that the samples are independent.

Sampling distribution of sum of statistics has mean $\mu_{S_1+S_2}$ and S.D. $\sigma_{S_1+S_2}$ given by

$$\mu_{S_1+S_2} = \mu_{S_1+S_2} \text{ and } \sigma_{S_1+S_2} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_1}^2}$$

1. Let $U_1 = (3, 7, 8)$, $U_2 = (2, 4)$ Find

a) μ_{U_1} (b) μ_{U_2} (c) Mean of the sampling distribution of the difference of means $\mu_{U_1-U_2}$

d) σ_{U_1} (e) σ_{U_2} (f) the standard deviation of the sampling distribution of the differences of means $\sigma_{U_1-U_2}$

Sol: Given $U_1 = \{3, 7, 8\}$ and $U_2 = \{2, 4\}$

$U_1 - U_2 = \{1, 5, 6, 4, 3, -1\}$

Now

(a) $\mu_{U_1} = \frac{3+7+8}{3} = 6$

(b) $\mu_{u_2} = \frac{2+4}{2} = 3$

(c) $\mu_{u_1 - u_2} = \frac{1+5+6+4+3-1}{6} = 3$

(d) $\sigma_{u_1} = \sqrt{\frac{(6-3)^2+(6-7)^2+(6-8)^2}{3}} = \sqrt{\frac{14}{3}}$

(e) $\sigma_{u_2} = \sqrt{\frac{(2-3)^2+(3-4)^2}{2}} = 1$

(f) $\sigma_{u_1 - u_2} = \sqrt{\frac{(1-3)^2+(5-3)^2+(6-3)^2+(4-3)^2+(3-3)^2+(-1-3)^2}{6}}$

$$= \sqrt{\frac{34}{6}} = \sqrt{\frac{17}{3}}.$$

SAMPLING DISTRIBUTION OF THE MEAN (σ UNKNOWN):-

For large sample of size ($n \geq 30$), even if standard deviation σ of Population is not known, it does not make any difference. Hence we can substitute the sample S.D. 's' in the place of σ and the sample S.D. 's', is calculated using the sample mean $\bar{x}$ by the formula

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

For small sample of size ($n<30$), when σ is unknown, it can be substituted by s. Provided we make the assumption that the sample is taken from normal population. we will discuss the t-distribution, F-distribution and χ^2- distribution in next chapter.

ESTIMATION

1

Estimate:- To find an unknown population parameter is a estimate statement.

Estimator:- The method of determining unknown population parameter is called estimator. For instance, sample mean is an estimator of population mean because sample mean is a method of determining the population mean.

A parameter can have one or two or many estimators.

Basically, there are two kinds of estimates to determine the statistic of the population parameter namely.

(a) Point estimation

(b) Interval estimation

① **Point estimation:-** If a single value is calculated as an estimate from an unknown population parameter. The procedure to find the parameter is called point estimation.

Properties of good estimator:- An estimator is said to be a good estimator if it is,

(i) unbiased (ii) consistend (iii) Efficient and sufficent.

(1) **Unbiased estimator:-** A statistic $\hat{\theta}$ is said to be an unbiased estimator if and only if the mean of the sampling distribution of estimator is equal to the parameter θ.

i.e, E [Statistic] = Parameter

$$E[\hat{\theta}] = \theta$$

(2) **Consistent:-** Any statistic $\bar{\theta}$ or $\hat{\theta}$ is said to be consistent if and only if, $E(\hat{\theta}) = \theta$ and $V(\hat{\theta}) \to 0$ as $n \to \infty$

(3) **Efficient:-** Suppose $\hat{\theta}_1, \hat{\theta}_2$ be two unbiased estimator of popln

Parameter θ and $V(\hat{\theta}_1)$ & $V(\hat{\theta}_2)$ be the variance of statistic. if $V(\hat{\theta}_1) < V(\hat{\theta}_2)$ then $\hat{\theta}_1$ is said to be efficient (or) more efficient unbiased estimator of θ.

Interval estimation:- In general, point estimator does not coincide with a true value of the parameter, so it is prefered to obtain arrange of values in an interval in which the parameter value lies.

Suppose α is the probability that the interval (a,b) does not include that the parameter θ.

Then $1-\alpha = P[a<\theta<b] = P[\theta, a, b]$

The interval (a,b) is called confidence interval $(1-\alpha)$ is called confidence coefficient and is generally given as $(1-\alpha)100\%$.. a & b are also known as confidence limits of parameter θ.

MAXIMUM ERROR OF ESTIMATE E FOR LARGE SAMPLES:-

The sample mean estimate very rarely equals to the mean of population μ, a point estimate is generally accompanied with a statement of error which gives difference between estimate and the quantity to be estimated. the estimator.

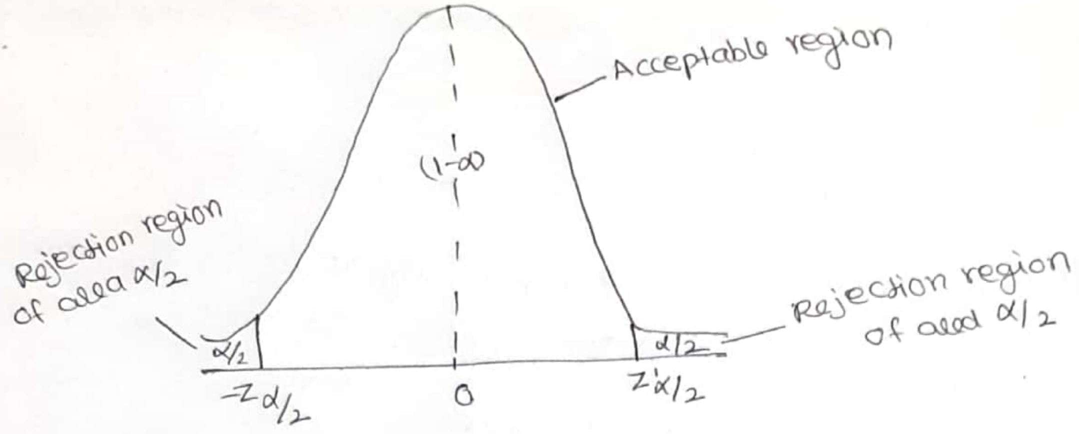
Thus error is $|\bar{x}-\mu|$

For large n, the random variable $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ is normal variate approximately

Then $P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$

where $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

Hence $P\left(-Z_{\alpha/2} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < Z_{\alpha/2}\right) = 1-\alpha$



$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha$$

Multiplying each term in the inequality by $\sigma/\sqrt{n}$ and then subtracting $\bar{x}$ from each term and multiplying -1.

$$\therefore P\left(-Z_{\alpha/2}(\sigma/\sqrt{n}) - \bar{x} < -\mu < Z_{\alpha/2}(\sigma/\sqrt{n}) - \bar{x}\right) = 1-\alpha$$

$$P(\bar{x} + Z_{\alpha/2}(\sigma/\sqrt{n}) \not> \mu \not> \bar{x} - Z_{\alpha/2}(\sigma/\sqrt{n})) = 1-\alpha$$

$$\therefore P(\bar{x} - Z_{\alpha/2}(\sigma/\sqrt{n}) < \mu < \bar{x} + Z_{\alpha/2}(\sigma/\sqrt{n})) = 1-\alpha.$$

Confidence interval for μ, σ known:-

If $\bar{x}$ is the mean of a random sample of size n from the population with known variance σ^2, $(1-\alpha)100\%$ confidence interval for μ is given by

$$\bar{x} - Z_{\alpha/2}(\sigma/\sqrt{n}) < \mu < \bar{x} + Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

where $Z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

So, the maximum error of estimate E with $(1-\alpha)$ probability is given by

$$\boxed{E = Z_{\alpha/2}(\sigma/\sqrt{n})}$$

Sample size:- When α, E, σ are known, the sample size n is given by

$$n = \left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2 \quad (\text{or}) \quad E_{max} = Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

when σ is unknown; σ is replace by s, s is the standard deviation of sample to determine E.

Thus the maximum Error estimate

$$E = t_{\alpha/2}\left(s/\sqrt{n}\right) \text{ with } (1-\alpha) \text{ probability}$$

Confidence interval μ, σ unknown:-

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal population with unknown variance σ^2, a $(1-\alpha)100\%$ confidence interval for μ is

$$\bar{x} - t_{\alpha/2}(s/\sqrt{n}) < \mu < \bar{x} + t_{\alpha/2}(s/\sqrt{n})$$

where $t_{\alpha/2}$ is the t-value with $\nu = n-1$ d.f., leaving an area of $\alpha/2$ to the right.

$\therefore$ The maximum error of estimate for small samples is given by

$$E = t_{\alpha/2}\, s/\sqrt{n-1}$$

where n = sample size, s = S.D. of means.

BAYESIAN ESTIMATION:-

Bayesian estimation is used to obtain mean and variance of posterior distribution of a population.

If the prior distribution on parameters mean μ_0 and variance σ_0^2 of a population are known then find the posterior distribution parameter of a given population. This

9

is called Bayesian estimation.

$$\mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2+\sigma^2} \quad \text{and} \quad \sigma_1 = \sqrt{\frac{\sigma_0^2\sigma^2}{n\sigma_0^2+\sigma^2}}$$

n = Sample size

$\bar{x}$ = sample mean

S = Standard deviation of sample use $S=\sigma$.

σ_0^2 = prior variance

σ^2 = sample variance

μ_0 = prior mean.

Here μ_1 and σ_1 are known as the mean and S.D. of the posterior distribution. In the computation of μ_1 and σ_1, σ^2 is assumed to be known, when σ^2 is unknown, which is generally the case, is replaced by sample variance s^2 provided $n \geq 30$ (large sample)

Bayesian interval for μ:

$(1-\alpha)100\%$ Bayesian interval for μ is given by

$$\mu_1 - Z_{\alpha/2}\cdot\sigma_1 < \mu < \mu_1 + Z_{\alpha/2}\cdot\sigma_1$$

Sufficient:- $T = t(x_1, x_2, \ldots x_n)$ is an estimator of a parameter θ, based on a sample $x_1, x_2 \ldots x_n$ of size n from the population with density $f(x,\theta)$. Such that the conditional distribution of $x_1, x_2 \ldots x_n$ given T is independent of θ, then T is a sufficient estimator of θ.

① In a study of an automobile insurance a random sample of 80 body repair costs had a mean of Rs. 472.36 and the S.D. of Rs 62.35. If $\bar{x}$ is used as a point estimate to the true average repair costs, with what confidence we can assert that the maximum error doesn't exceed Rs. 10?

Sol. Size of random sample $n=80$.

The mean of random sample $\bar{x}$ = Rs 472.36

Standard deviation, σ = Rs 62.35

Maximum error of estimate, E_{max} = Rs 10

we have $E_{max} = Z_{\alpha/2} \cdot \sigma/\sqrt{n}$

$$\Rightarrow Z_{\alpha/2} = \frac{E_{max} \cdot \sqrt{n}}{\sigma} = \frac{10\sqrt{80}}{62.35}$$

$$\Rightarrow Z_{\alpha_2} = 0.9236$$

$$\Rightarrow 1-\alpha/2 = Z_{\alpha/2} = 0.9236 \Rightarrow \alpha = 0.1528$$

$\therefore$ confidence $= (1-\alpha)100\% = 84.72\%$.

② If we can asseet with 95% that the maximum error is 0.05 and P=0.2 find the size of the sample.

Sol: Given P = 0.2, E = 0.05

we know that maximum error $E = Z_{\alpha/2}\sqrt{\frac{PQ}{n}}$

$$\Rightarrow 0.05 = 1.96\sqrt{\frac{0.2 \times 0.8}{n}}$$

$$\Rightarrow n = \frac{0.2 \times 0.8 \times (1.96)^2}{(0.05)^2}$$

$$n = 246$$

③ Assuming that $\sigma = 20.0$, how laege a random sample be taken to asseet with probability 0.95 that the sample mean will not diffel from the true mean by more than 3.0 points?

Sol: Given maximum error E = 3.0 and $\sigma = 20.0$

we have $Z_{\alpha/2} = 1.96$

W.K.T. $n = \left(\frac{Z_{\alpha/2}\,\sigma}{E}\right)^2$

$$n = \left(\frac{1.96 \times 20}{3}\right)^2 = 170.74$$

$$n \cong 171.$$

4. What is the maximum error one can expect to make with probability 0.90, when using the mean of a random sample of size $n=64$ to estimate the mean of population with $\sigma^2 = 2.56$

10

Sol: Here $n = 64$

The probability $= 0.90$

$\sigma^2 = 2.56 \Rightarrow \sigma = \sqrt{2.56} = 1.6$

Confidence limit $= 90\%$.

$\therefore Z_{\alpha/2} = 1.645$

Hence maximum error $E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$= 1.645 \times \frac{1.6}{\sqrt{64}} = 0.329.$

5. A random sample of size 100 has a standard deviation of 5. what can you say about the maximum error with 95% confidence.

Sol: Given $s = 5$, $n = 100$

$Z_{\alpha/2}$ for 95% confidence $= 1.96$

W.K.T. Maximum error $E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$\therefore E = (1.96) \frac{5}{\sqrt{100}} = 0.98.$

6. A random sample of 400 items is found to have mean 82 and S.D. of 18. Find the maximum error of estimation at 95% confidence interval. Find the confidence limits for the mean if $\bar{x} = 82$.

Sol: Give S.D. $\sigma = 18$

$n = 400$

$Z_{\alpha/2}$ for 95% confidence $= 1.96$

Sample mean $= \bar{x} = 82$

Maximum error, $E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$= \frac{1.96 \times 18}{20} = 1.764$

The limits for the confidence are

$$\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$\therefore$ Confidence limits are 80.236 and 83.764.

(7) A sample of size 300 was taken whose variance is 225 and mean 54. Construct 95% confidence interval for the mean.

Sol: Since the sample size 300 is large (>30), normal distribution is used as the sampling distribution.

Here $n = 300$, $\bar{x}$ = Sample mean = 54, $\sigma = \sqrt{225} = 15$

$\therefore$ S.E. of $\bar{x} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{300}} = 0.866$

95% confidence limits for the population mean are.

$\bar{x} \pm 1.96$ (S.E. of $\bar{x}$) $= 54 \pm 1.96(0.866)$

$= 55.697$ & 52.3

$\therefore$ The required confidence interval is (55.697, 52.3)

(8) A population random variable has mean 100 and S.D. 16 what are the mean and S.D. of the sample mean for the random sample of size 4 drawn with replacement.

Sol. Given $\mu = 100$, $\sigma = 16$, $n = 4$

Since the sampling is done with replacements, the population may be considered as infinite.

we have to find $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$.

$\therefore \mu_{\bar{x}} = \mu = 100$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{4}} = 8$

(9) Find 95% confidence limits for the mean of a normality distri-buted population from which the following sample was taken.

15, 17, 10, 18, 16, 9, 7, 11, 13, 14.

(10) A random sample of size 81 was taken whose variance is 20.25 and mean is 32, construct 98% confidence interval.

TEST OF HYPOTHESIS

Test of hypothesis:- When parameteic values are unknown, we estimate then through sample values. But the problem arises when the sample provides a value, which is either exactly equal to the parameteic value, not too far, In that situation one has to develop some procedure which enables one to decide whether to accept a value or not on the basis of sample values, such a procedure known as testing of hypothesis.

In many circumstances, we are to make decisions about population on the basis of only sample information. For example, on the basis of sample data

(i) a drug chemist is to decide whether a new drug is really effective in curing a disease.

(ii) a quality control manager is to determine whether a process is working properly.

TEST OF STATISTICAL HYPOTHESIS:-

A statistical test of hypothesis is a rule or procedure which makes one to decide about the acceptance or rejection of the hypothesis it can be denoted by H.

Hypothesis are two types

(i) Null Hypothesis

(ii) Alternative Hypothesis.

1) Null Hypothesis:- For applying the tests of significance, we irst setup a hypothesis which is a statement about the pulation parameter. Such a hypothesis is useually hypothesy

no difference is called a null hypothesis.

i.e., There is no significance difference between two parameter.

and it is denoted by H_0.

<u>Alternative Hypothesis</u>:- Any hypothesis which is complementary the null hypothesis is called an alternative hypothesis.

and it is denoted by H_1.

i.e., There is a significance difference b/w two parameter.

<u>Critical region</u>:- Let $x_1, x_2 \dots x_n$ be the sample observations and S be the sample space. we devide the whole sample space 'S' into two disjoint parts w and S-w. A region in the sample space S. which amounts to rejection to H_0 is called critical region (or) region of rejection.

w is the critical region and S-w is the acceptance region. If the sample points falls in the subset W. H_0 is rejected otherwise H_0 is accepted.

<u>ERRORS IN SAMPLING</u>

There are two types of errors in sampling

① Type I error

② Type II error.

Type I error (or) first kind of error :-

Reject H_0 when H_0 is true. it is also known as rejection error. Probability of type I error it is denoted by α.

i.e., $\alpha = P[\text{Reject } H_0 / H_0 \text{ is true}]$.

Type II error (OR) Second kind of error :-

Accept H_0 when H_0 is false. It is also known as acceptance error. Probability of type-II error is denoted by β.

i.e., $\beta = P[\text{accept } H_0 / H_0 \text{ is false}]$

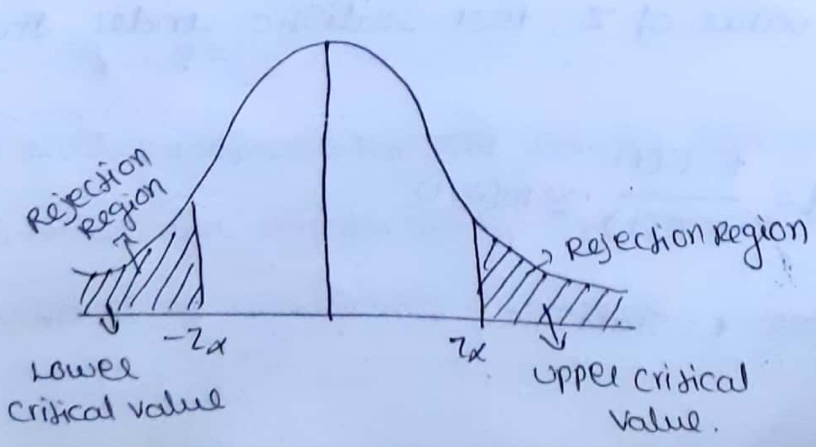
Type II error more serious than type I error.

Level of significance :- The probability of type I error is known as level of significance. The l.o.s. usually employed in testing of hypothesis are 5% and 1%, α is always fixed in advance before collecting the sample information.

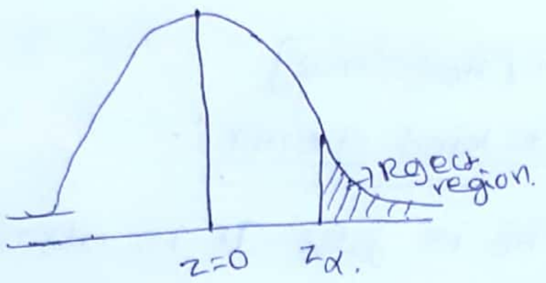
TWO-TAILED AND ONE-TAILED TESTS :-

Hypothesis is such that it leads two alternatives to the H_0, it is said to be a two-tailed test.

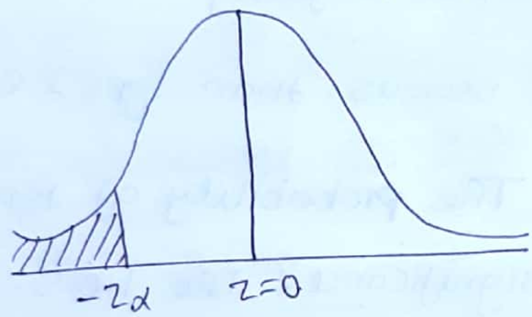
i.e., $H_1: \mu \neq \mu_0$ $[\mu > \mu_0 \text{ or } \mu < \mu_0]$



Right tailed test:- If alternative hypothesis $H_1: \mu > \mu_0$ to the H_0 it is said to be a right tailed test. In this situation the half of the area of critical region lies half on the right.



Left tailed test:- If $H_1: \mu < \mu_0$ to the H_0 it is known as left tailed test.



PROCEDURE FOR TESTING OF HYPOTHESIS:

① setup the Null hypothesis H_0: There is no significant difference b/w statistic and parameter

② Setup alternative hypothesis H_1.

③ choose the appropriate level of significance α.

(1% or 5% or any percentage)

④ Calculate the value of Z, test statistic under the null hypothesis.

$$|Z| = \frac{t - E(t)}{S.E(t)} \sim N(0,1)$$

where t = statistic.

⑤ Conclusion:- Compare calculated value of z with the tabulated value at α l.o.s.

If cal. value $z <$ table value of z, then we accept the null hypothesis at α l.o.s. and otherwise reject H_0.

Critical value of z	l.o.s.		
z_α	1%	5%	10%
Two tailed test	2.58	1.96	1.645
Right tailed test	2.33	1.645	1.28
left tailed test	-2.33	-1.645	-1.28

TEST OF SIGNIFICANCE FOR LARGE SAMPLES:-

Standard normal variate $z = \frac{x-\mu}{\sigma}$

$\mu = np$

$\sigma = \sqrt{npq}$

① A coin was tossed 960 times and returned heads 183 times. Test the hypothesis that the coin is unbiased. Use a 0.05 level of significance.

Sol: Here $n = 960$, $x = 183$

P = probability of getting head $= \frac{1}{2}$

$\therefore q = 1 - p = \frac{1}{2}$

1) Null Hypothesis H_0: The coin is unbiased

2) Alternative Hypothesis H_1: The coin is based

3) Level of significance; $\alpha = 0.05$

4. The test statistic is :

Standard normal variate $z = \frac{x-\mu}{\sigma}$

$p = \frac{1}{2}, \quad q = \frac{1}{2}$

$n = 960$

$\mu = np = 960(\frac{1}{2}) = 480$

$\sigma = \sqrt{960(\frac{1}{2})(\frac{1}{2})}$

$= 15.49$

$z = \frac{x-\mu}{\sigma}$

$= \frac{183-480}{15.49}$

$= \frac{-297}{15.49} = -19.17$

$\therefore |z| = 19.17$

At $z_{cal} \not< z_{tab}$

$\therefore$ Ho is rejected at 5% l.o.s.

$\therefore$ The coin is biased.

(2) A die is tossed 960 times and it falls with 5 upwards 184 times. Is the die unbiased at a level of significance of 0.01?

As:- accept

(3) A coin was tossed 400 times and returned heads 216 times. Test the hypothesis that the coin is unbiased. Use a 0.05 L.O.S.

As: accept.

LARGE SAMPLE TESTS

Under large sample tests, we will see four important tests to test the significance.

1) Testing of significance for single Proportion
2) " " for difference of Proportion
3) " " for single mean
4) " " for difference of means.

TEST OF SIGNIFICANCE OF A SINGLE MEAN :-

Let a random sample of size $n (n \geq 30)$ has the sample mean $\bar{x}$, and μ be the population mean. Also the pop mean μ has a specified value μ_0.

The test statistic $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

we use $s = \sigma$

NOTE:-

① 95% confidence limits $\bar{x} \pm 1.96\, \sigma/\sqrt{n}$

② 99% confidence limits $\bar{x} \pm 2.58\, \sigma/\sqrt{n}$

③ 98% confidence limits $\bar{x} \pm 2.33\, \sigma/\sqrt{n}$

① An oceanographer wants to check whether the depth of the ocean in a certain region is 57.4 fathoms, as had previously been re-corded. What can be conclude at the 0.05 level of significance, if readings taken at 40 random locations in the given region yielded a mean of 59.1 fathoms with a standard deviation of 5.2 fathoms.

Sol: Given $n = 40$, $\bar{x} = 59.1$ and $\sigma = 5.2$

1. Null Hypothesis H_0: $\mu = 57.4$

2. Alternative hypothesis H_1: $\mu \neq 57.4$

3) Level of significance $= 0.05$

4) The test statistic is $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

$$= \frac{59.1-57.4}{5.2/\sqrt{40}} = 2.067$$

$$\cancel{= \frac{1.2}{0.7875} = 1.5238}$$

Tabulated value of z at 5% level of significance is 1.96

Hence calculated $z >$ tabulated z.

$\therefore$ The null hypothesis H_0 is rejected.

② A sample of 900 members has a mean of 3.4 cms and S.D 2.61 cms. Is the sample from a large population of mean 3.25 cms and S.D 2.61 cms. If the population is normal and its mean is unknown find the 95% fiducial limits of true mean.

Sol Given $n = 900$ $\mu = 3.25$

$\bar{x} = 3.4$ cm $\sigma = 2.61$

and $S = 2.61$

1) Null Hypothesis H_0: Assume that the sample has been drawn from the population with mean $\mu = 3.25$.

2) Alternative Hypothesis H_1: $\mu \neq 3.25$

3) The test statistic is, $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

$$= \frac{3.4-3.25}{2.61/\sqrt{900}}$$

$$= 1.724$$

i.e., $Z = 1.724 < 1.96$

$\therefore$ we accept the null hypothesis H_0.

Scanned with CamScanner

i.e., The sample has been drawn from the population with

mean $\mu = 3.25$.

95% confidence limits are given by

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}} = 3.4 \pm 1.96 \times \frac{2.61}{\sqrt{900}}$$

$$= 3.4 \pm 0.1705$$

i.e., 3.57 and 3.2295

③ A sample of 400 items is taken from a population whose S.D. is 10. The mean of the sample is 40. Test whether the sample has come from a population with mean 38. Also calculate 95% confidence interval for the population.

Sol: Given $n = 400$, $\bar{x} = 40$, $\mu = 38$ and $\sigma = 10$

1. Null hypothesis $H_0: \mu = 38$

2. Alternative hypothesis $H_1: \mu \neq 38$

3. Level of significance, $\alpha = 0.05$

4. The test statistic is $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$$= \frac{40 - 38}{10/\sqrt{400}} = 4$$

i.e, $Z = 4 > 1.96$

$\therefore$ we reject the null hypothesis H_0

i.e., the sample is not from the population whose mean is 38.

95% confidence interval is $\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$

i.e., $\left(40 - \frac{1.96(10)}{\sqrt{400}}, 40 + \frac{1.96(10)}{\sqrt{400}}\right)$

$= \left(40 - \frac{1.96 \times 10}{20}, 40 + \frac{1.96 \times 10}{20}\right)$

$= (39.02, 40.98)$

4) In a random sample of 60 workers, the average time taken by them to get to work is 33.8 minutes with a standard deviation of 6.1 minutes. Can we reject the null hypothesis $\mu = 32.6$ minutes in favour of alternative null hypothesis $\mu > 32.6$ at $\alpha = 0.025$ l.o.s.

Sol: Given $n = 60$, $\bar{x} = 33.8$, $\mu = 32.6$ and $\sigma = 6.1$

1. Null hypothesis $H_0: \mu = 32.6$.

2. Alternative hypothesis $H_1: \mu > 32.6$

3. Level of significance $\alpha = 0.025$

4. The test statistic is $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$= \frac{33.8 - 32.6}{6.1/\sqrt{60}} = \frac{1.2}{0.7875}$

$= 1.5238$

Tabulated value of Z at 0.025 l.o.s is 2.58

Hence calculated Z < tabulated Z

$\therefore$ The null hypothesis H_0 is ~~rejected~~ accepted.

5) A sample of 64 students have a mean weight of 70 kgs. Can this be regarded as a sample from a population with mean weight 65 kgs and standard deviation 25 kgs.

6) The mean life time of a sample of 100 light tubes produced by a company is found to be 1560 hrs with a population S.D. of 90 hrs. Test the hypothesis for $\alpha = 0.05$ that the mean life time of the tubes produced by the company is 1580 hrs.

TEST FOR EQUALITY OF TWO MEANS

(TEST OF SIGNIFICANE FOR DIFFERENCE OF MEANS OF TWO LARGE SAMPLES)

Let $\bar{x}_1$ and $\bar{x}_2$ be the sample means of two independent large random samples sizes n_1 and n_2 drawn from two populations having mean μ_1 and μ_2 and standard deviations σ_1 and σ_2. To test whether the population means are equal.

The test Statistic is

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

where $\delta = \mu_1 - \mu_2$ = given constants

$H_0: \mu_1 = \mu_2$, then the test statistic

becomes $$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

use $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

σ is not known we can use an estime of

$$\sigma^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}$$

① The means of two large samples of sizes 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from the same population of S.D 2.5 inches.

Sol: Let μ_1 and μ_2 be the means of the two populations.

Given $n_1 = 1000$, $n_2 = 2000$

and $\bar{x}_1 = 67.5$ inches , $\bar{x}_2 = 68$ inches.

population S.D. $\sigma = 2.5$ inches.

1. Null Hypothesis H_0: The sample have been draw from the Same population of S.D. 2.5 inches.

i.e., $\mu_1 = \mu_2$ and $\sigma = 2.5$ inches.

2. Alternative hypothesis $H_1: \mu_1 \neq \mu_2$

3. The test statistic is $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma^2}{n_1} + \dfrac{\sigma^2}{n_2}}}$

$$= \frac{67.5 - 68}{\sqrt{(2.5)^2\left(\frac{1}{1000} + \frac{1}{2000}\right)}}$$

$$\Rightarrow Z = \frac{-0.5}{0.0968} = -5.16$$

$\therefore |Z| = 5.16 > 1.96$

i.e., the calculated value of $Z >$ the table value of Z.

Hence the null hypothesis H_0 is rejected at 5% l.o.s. and we conclude that the samples are not drawn from the same population of S.D. 2.5 inches.

② Samples of students were drawn from two universities and from their weights in kilograms, mean and standard deviations are calculated and show below. Make a large sample test to test the significance of the difference between the means.

	Mean	S.D.	Size of the sample
University A	55	10	400
University B	57	15	100.

Sol: Given $\bar{x}_1 = 55$, $\bar{x}_2 = 57$, $n_1 = 400$, $n_2 = 100$.

$S_1 = 10$ and $S_2 = 15$

1. Null Hypothesise H_0: $\bar{x}_1 = \bar{x}_2$ i.e., there is no difference
2. Alternative Hypothesis H_1: $\bar{x}_1 \neq \bar{x}_2$
3. Level of significance, $\alpha = 0.05$
4. Critical region : Accept H_0 if $-1.96 < Z < 1.96$
5. The test statistic is

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$= \frac{55-57}{\sqrt{\frac{100}{400} + \frac{225}{100}}} = \frac{-2}{\sqrt{\frac{1}{4} + \frac{9}{4}}}$$

$$= -1.26.$$

$\therefore |Z| = 1.26 < 1.96$

Hence, we accept the Null Hypothesis H_0 at 5% l.o.s. and conclude that there is no significant difference between the means.

(3) A researcher want to know the intelligence of students in a school. He selected two groups of students. In the first group there 150 students having mean IQ of 75 with a S.D. of 15 in the second group there are 250 Students having mean IQ of 70 with S.D. is 20.

(4) A simple Sample of the height of 6400 Englishmen has a mean of 67.85 inches and a S.D. of 2.56 inches while a simple sample of heights of 1600 Austrians has a mean of 68.55 inches and S.D. of 2.52 inches. Do the data indicate the Austrains are on the average ~~of~~ taller than the Englishmen? (use α is 0.01)

TEST OF SIGNIFICANCE FOR SINGLE PROPORTION:-

Suppose a large random sample of size n ~~has a sample proportion p of members~~ is taken from a normal population. To test the significance difference between the sample proportion p and the population P, we use the test statistic

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$ where n is the sample size.

NOTE ①:- Limits for population proportion P are given by $p \pm 3\sqrt{\frac{pq}{n}}$

where $q = 1 - p$

② Confidence interval for proportion P of large sample at α l.o.s. is

$$p - z_{\alpha/2}\sqrt{\frac{PQ}{n}} < P < p + z_{\alpha/2}\sqrt{\frac{PQ}{n}}$$ where $Q = 1 - P$ and

$z_{\alpha/2} = 1.96$ (for 95%)

$z_{\alpha/2} = 2.33$ (for 98%)

$z_{\alpha/2} = 2.58$ (for 99%)

① A manufacturer claimed that atleast 95% of the equipment which he supplied to a factory to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at 5% l.o.s.

Sol. Given sample size $n = 200$

Number of pieces confirming to specification $= 200 - 18 = 182$

$\therefore p =$ proportion of pieces confirming to specifications.

$$= \frac{182}{200} = 0.91$$

P= population proportion = $\frac{95}{100}=0.95$.

1. Null Hypothesis H_0: The proportion of pieces confirming to specifications.

i.e., P=95%.

2. Alternative Hypothesis H_1: $P<0.95$ (left-tail test)

3. The test statistic is

$$Z=\frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.91-0.95}{\sqrt{\frac{0.95\times0.05}{200}}}$$

$$Z=\frac{-0.04}{0.0154}=-2.59$$

Since alternative hypothesis is left tailed, the tabulated value of z at 5% l.o.s. is 1.645.

Since calculated value of $|z|=2.6$ is greater than 1.645, we reject the Null Hypothesis H_0 at 5% l.o.s. and Conclude that the manufacture's claim is rejected.

② 20 people were attacked by a disease and only 18 survived. will you reject the hypothesis that the survival rate if attacked by this disease is 85% in favour of the hypothesis that is more at 5% l.o.s?

Sol:

n = Sample size = 20

X = number of survived sample = 18

p = proportion of survived people = $\frac{X}{n}=\frac{18}{20}=0.9$

P = 0.85

$\therefore Q=1-P=1-0.85$

$=0.15$

1. Null Hypothesis H_0: P=0.85

2. Alternative " H_1: P>0.85 (Right tailed)

3. Level of significance, $\alpha = 0.5$

4. The test statistic is.

$$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.9-0.85}{\sqrt{\frac{0.85 \times 0.15}{20}}} = \frac{0.5}{0.8} = 0.625$$

$\therefore$ Calculated $z = 0.625$

Tabulated z at 5% l.o.s. $Z_\alpha = 1.645$

Since calculated $z <$ tabulated z,

we accept the N.H. H_0.

i.e., The proportion of the survived people is 0.85.

③ A random sample of 500 pineapples was taken from a lage consignment and 65 were found to be bad. Find the percentage of bad pineapples in the consignment.

Sol. $n = 500$

p = Proprotion of bad pineapples in the sample

$$= \frac{65}{500} = 0.13$$

$q = 1-p = 0.87$

we know that the limits for population proportion P are given by

$$p \pm 3\sqrt{\frac{pq}{n}} = 0.13 \pm 3\sqrt{\frac{0.13 \times 0.87}{500}} = 0.13 \pm 0.045$$

$$= (0.085, 0.175)$$

$\therefore$ The percentage of bad pineapples in the consignment lies between 8.5 and 17.5.

④ In a random sample of 160 workers exposed to a certain amount of radiation, 24 experienced some ill effects. construct a 99% con-fidence interval for the corresponding true percentage.

20

sol: we have $x = 24$

$n = 160$

$p = \frac{24}{160} = 0.15, \quad Q = 1 - P = 0.85$

Now $\sqrt{\frac{PQ}{n}} = \sqrt{\frac{0.15 \times 0.85}{160}} = 0.028$

Confidence interval at 99% l.o.s. is

$$\left(P - 3\sqrt{\frac{PQ}{n}}, \; P + 3\sqrt{\frac{PQ}{n}}\right)$$

i.e., $(0.15 - 3 \times 0.028, \; 0.15 + 3 \times 0.028)$

$(0.065, 0.234)$

⑤ In a sample of 500 from a village in Andra pradesh, 280 are found to be rice eaters and the rest wheat eaters. Can we assume that the both articles are equally popular.

⑥ In a hospital 480 feamales and 520 male babies were born in a week. Do these figure confirm the hypothesis that males and females are born in equal number?

TEST FOR EQUALITY OF TWO PROPORTIONS (OR TEST OF SIGNIFICANCE OF DIFFERENCE BETWEEN TWO SAMPLE PROPORTIONS - LARGE SAMPLES)

Let p_1 and p_2 be the proportions in two large random samples of sizes n_1 and n_2 drawn from two populations having P_1 and P_2.

$\therefore$ Standard Error of Difference = $S.E.(P_1 - P_2) = \sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$

a) When the population proportions P_1 and P_2 are known.

Hence the test statistic is

$$Z = \frac{P_1 - P_2}{S.E.(P_1 - P_2)} = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

When the population proportions P_1 and P_2 are known but sample proportions p_1 and p_2 are known.

Hence the test statistic is

$$Z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ so that $q = 1 - p$.

① Random samples of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men and 325 women were in favour of the proposal. Test the hypothesis that proportions of men and women in favour of the proposal are same, at 5% level?

Sol: Given sample size $n_1 = 400$, $n_2 = 600$

Proportion of mean $p_1 = \frac{200}{400} = 0.5$

Proportion of women $p_2 = \frac{325}{600} = 0.541$

1) Null hypothesis H_0: Assume that there is no significant difference between the option of men and women as far as proposal of flyover is concerned.

i.e., $H_0: P_1 = P_2 = P$

2) Alternative hypothesis $H_1: P_1 \neq P_2$ (two tailed)

3. The test statistic is

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

where $p = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

$$= \frac{400 \times \frac{200}{400} + 600 \times \frac{325}{600}}{400+600} = \frac{525}{1000} = 0.525$$

and $q = 1 - p = 1 - 0.525$
$= 0.475$

$$\therefore z = \frac{0.5 - 0.541}{\sqrt{0.525 \times 0.425\left(\frac{1}{400}+\frac{1}{600}\right)}}$$

$$= \frac{-0.041}{0.032} = -1.28$$

$|z| = 1.28$

since $|z| < 1.96$, we accept the null hypothesis H_0 at 5% l.o.s.

i.e., there is no difference of opinion between men and women as far as proposal of flyover is concerned.

② In two large populations, there are 30% and 25% respectively of fair haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations.

sol: Given $n_1 = 1200$, $n_2 = 900$

P_1 = Proportion of fair haired people in the first population.

$$P_1 = \frac{30}{100} = 0.3.$$

P_2 = Proportion of fair haired people in the second population

$$P_2 = \frac{25}{100} = 0.25.$$

1. Null Hypothesis H_0: Assume that the sample proportions are equal i.e., the difference in population proportions is likely to be hidden in sampling.

i.e., $H_0: P_1 = P_2$

2. Alternative Hypothesis: $H_1: P_1 \neq P_2$

3. The test statistic is $Z = \dfrac{P_1 - P_2}{\sqrt{\dfrac{P_1 Q_1}{n_1} + \dfrac{P_2 Q_2}{n_2}}}$

where $Q_1 = 1 - P_1 = 1 - 0.3 = 0.7$

$Q_2 = 1 - P_2 = 1 - 0.25 = 0.75$

$$\therefore Z = \frac{0.3 - 0.25}{\sqrt{\frac{0.3 \times 0.7}{1200} + \frac{0.25 \times 0.75}{900}}}$$

$= 2.55$

i.e., $Z = 2.55$

Since $Z > 1.96$, therefore we reject the Null Hypothesis H_0 at 5% l.o.s. i.e., the sample proportions are not equal. Thus we conclude that the difference in population proportions is unlikely that the real difference will be hidden

(3) In a random sample of 1000 persons from town A, 400 are found to be consumers of wheat. In a sample of 800 from town B, 400 are found to be consumers of wheat. Do these data reveal a significant difference between town A and town B, so far as the proportion of wheat is concerned.

(4) A cigarette manufacturing firm claims that its brand A line of cigarettes outsells its brand B by 8%. If it is found that 42 out of a sample of 200 smokers prefer brand A and 18 out of another sample of 100 smokers prefer brand B. test whether the 8% difference is valid claim.

SMALL SAMPLES

t-test:- If a sample with less than 30 items then we use t-test for testing of hypothesis. In this we estimate the population variance from the sample values.

i.e., If $x_1, x_2, \ldots x_n$ are the values in the sample with sample mean $\bar{x}$ and standard deviation S drawn from a population with mean μ and S.D. σ.

$$S^2 = \frac{1}{n-1}\sum(x_i-\bar{x})^2 \text{, then}$$

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$$

is a random variable having the t-distribution with n-1 d.f. and with p.d.f.

$$f(t) = y_0\left(1+\frac{t^2}{v}\right)^{-\frac{v+1}{2}}$$ where $v = n-1$ and y_0 is a constant.

APPLICATIONS OF THE t-DISTRIBUTION:-

The t-distribution has a wide number of applications in statistics, some of them are given below:

1. To test the significance of the sample mean, when population variance is not given.
2. To test the significance of the mean of the sample. i.e., to test if the sample mean differs significantly from the population mean.
3. To test the significance of the difference between two sample means or to compare two samples
4. To test the significance of an observed sample correlation coefficient and sample regression coefficient.

① A random sample of size 25 from a normal population has the mean $\bar{x} = 47.5$ and the standard deviation $s = 8.4$. Does this information tend to support or refuse the claim that the mean of the population is $\mu = 42.5$?

Sol: Given n = The size of the sample = 25

$\bar{x}$ = The mean of the sample = 47.5

μ = The population mean = 42.5

S = S.D of sample = 8.4

we have t-distribution

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{47.5-42.5}{8.4/\sqrt{25}}$$

$$= \frac{5\sqrt{25}}{8.4} = 2.98$$

This value of t has 24 d.f. ($\because \nu = n-1 = 25-1 = 24$)

From the table of t-distribution for $\nu = 24$ with $\alpha = 0.005$ is 2.797. we conclude that the information given in the data of this example tend to refuse the claim that the mean of the population is $\mu = 42.5$ (i.e, μ cannot be 42.5)

F-test:- F-test can be used for test of equality of population variances.

Suppose we want to test whether two independent samples x_i ($i = 1, 2, \dots n_1$), y_j ($1, 2 \dots n_2$) has been drawn from the normal population with same variances σ^2, we consider $H_0: \sigma_x^2 = \sigma_y^2 = \sigma^2$.

i.e., Population variances are equal.

To test above H_0, the test statistic is given by

$$F = \frac{S_1^2}{S_2^2} \sim F((n_1-1), (n_2-1)) \text{ d.f.}$$

23

where $S_1^2 = \frac{1}{n_1-1}\sum(x_i-\bar{x})^2$

$$S_2^2 = \frac{1}{n_2-1}\sum(y_i-\bar{y})^2$$

Conclusion:- Compare calculated F with $F_{(n_1-1),(n_2-1)}$ d.f. at certain l.o.s. (5% or 1%.)

i) $F_{cal} < F_{tab}$, accept H_0

$F_{cal} > F_{tab}$ Reject H_0.

χ^2-distribution:- Let s^2 be the sample variance of size n, taken from a normal population having the variance σ^2.

Then $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{\sigma^2}$ $\left[\because S^2 = \sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{n-1}\right]$

is the value of a random variable having the χ^2-distribution with $\nu = n-1$ d.f.

APPLICATIONS OF χ^2 DISTRIBUTION:-

1. To test the goodness of fit.
2. To test the independence of attributes
3. To test the homogeneity of independent estimation of the population variance.
4. To test the homogeneity of independent estimation of the population correlation coefficient.

① For an F-distribution, find

(a) $F_{0.05}$ with $\nu_1 = 7$ and $\nu_2 = 15$

(b) $F_{0.01}$ with $\nu_1 = 24$ and $\nu_2 = 19$

c) $F_{0.95}$ width $\nu_1 = 19$ and $\nu_2 = 24$

d) $F_{0.99}$ width $\nu_1 = 28$ and $\nu_2 = 12$

sol:

a) From table $F_{0.05}$ with $\nu_1 = 7$ and $\nu_2 = 15$ is 2.71

b) $F_{0.01}$ with $\nu_1 = 24, \nu_2 = 19$ is 2.92

c) $F_{0.95}(19,24) = \dfrac{1}{F_{0.05}(24,19)} = \dfrac{1}{2.11} = 0.473$

d) $F_{0.99}(28,12) = \dfrac{1}{F_{0.01}(12,28)} = \dfrac{1}{2.90} = 0.34482$

TEST OF SIGNIFICANCE FOR SMALL SAMPLES :-

There are three important tests for, that is

(i) Students 't' Test.

(ii) F-test

(iii) χ^2-test.

① Students t-test: single mean

The Students t is defined by the test

Statistic $\quad t = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n-1}}$

$\bar{x}$ = mean of a sample

n = size of the sample

σ = S.D. of the sample

μ = mean of the population supposed to be normal

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

① The average breaking strength of the steel rods is specified to be 18.5 thousand pounds. To test this sample of 14 rods were tested. The mean and standard deviations obtained were 17.85 and 1.955 respectively. Is the result of experi-ment significant?

Sol: Given sample size $n = 14$

sample mean $\bar{x} = 17.85$

S.D. (S) $= 1.955$

population mean, $\mu = 18.5$

Degrees of freedom $= n-1 = 13$

1. Null Hypothesis H_0: The resut of the expeeiment is not significant.

2. Alternative Hypothesis H_1: $\mu \neq 18.5$

3. Level of significance $\alpha = 0.05$

4. The test statistic is

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n-1}}$$

$$= \frac{17.85 - 18.5}{1.955/\sqrt{13}} = \frac{0.65}{0.542} = -1.199$$

$\therefore |t| = 1.199$

i.e., calculated $t = 1.199$

Tabulated t at 5% l.o.s. for 13 d.f. for two tailed test $= 2.16$

Since calculated value $t <$ table value t, we accept the Null Hypothesis H_0 at 5% level and conclude that the result of the expeeiment is not significant.

② A random sample of six steel beams has a mean compressive strength of 58,392 P.S.i (Pounds per square inch) with a S.D. of 648 p.s.i. use this information and l.o.s $\alpha = 0.05$ to test whether the true avg compressive strength of the steel from which they sample came is 58,000 P.S.i. Assume normality.

Sol: we have

n = sample size (number of steel beams) $= 6 < 30$

∴ The sample is small.

$\bar{x}$ = Sample mean (average compressive strength) = 58392 psi.

S = S.D. of six Beams = 648 psi

Degrees of freedom $= n-1 = 6-1 = 5$

In this problem σ is known and $n<30$. Hence we use t-distribution.

1. Null Hypothesis: $H_0: \mu = 58000$

2. Alternative Hypothesis $H_1: \mu \neq 58000$

3. l.o.s. $\alpha = 0.05$

4. Critical region: since A.H. is of the type '≠' the test is two-tailed and the critical region is $-3.365 < t < 3.365$

5. The test statistic is $t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$

$$= \frac{58392-58000}{648/\sqrt{5}}$$

$$= 1.353$$

since $t = 1.353 < 3.365 = t_{\alpha/2}$, we accept the null hypothesis H_0.

Hence the average compressive strength of the steel beam is not equal to 58000 psi.

(3) A sample of 100 iron bars is said to be drawn from a large number of bars whose lengths are normally distributed with mean 4 feet and S.D. 6ft. If the sample mean is 4.2 feet. Can the sample be regarded as a truly random sample?

(4) A sample of 155 members has a mean 67 and S.D. 5.2. In this sample has been taken from a large population of mean 70.

Problems related to student's t-Test (when S.D. of the sample is not given directly)

(1) A random sample of 10 boys had the following I.Q's: 70, 120, 110, 101, 88, 83, 95, 98, 107 and 100 (a) Do these data support the assumption of a population mean I.Q. of 100? (b) Find a reasonable range in which most of the mean I.Q. values of samples of 10 boys lie.

Sol: (a) Here S.D. and mean of the sample is not given directly. we have to determine these S.D. and mean as follows.

Mean $\bar{x} = \frac{\Sigma x}{n} = \frac{972}{10} = 97.2.$

x	$x-\bar{x}$	$(x-\bar{x})^2$
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
972		1833.60

we know that $s^2 = \frac{1}{n-1}\sum_i (x_i-\bar{x})^2$

$$= \frac{1833.60}{9}$$

$\therefore$ Standard deviation $s = \sqrt{203.73}$

$$= 14.27$$

1. Null Hypothesis H_0: The data support the assumption of a population mean I.Q. of 100 in the population

2. Alternative Hypothesis: $H_1: \mu \neq 100$ (two-tailed test)

3. L.O.S. $\alpha = 0.05$

4. The test statistic is

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{97.2-100}{14.72/\sqrt{10}} = -0.62$$

$\therefore |t| = 0.62$ i.e., calculated value of $t = 0.62$

Tab. value of t for (10-1) d.f. i.e., 9 d.f. at 5% l.o.S is 2.26.

Since cal. value t > tab value t. we accept the null hypothesis H_0, i.e., the data support the assumption of mean I.Q. of 100 in the population.

(b) The 95% confidence limits are given by $\bar{x} \pm t_{0.05} s/\sqrt{n}$

$$= 97.2 \pm 2.26 \times 4.512 = 107.4 \text{ and } 87$$

$\therefore$ The 95% confidence limits within which the mean I.Q. values of sample of 10 boys will be (87, 107.40)

② The heights of 10 males of a given locality are found to be 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches. Is it reasonable to believe that the avg height is greater than 64 inches? Test at 5% l.o.S. assuming that for 9 d.f. ($t = 1.833$ at $\alpha = 0.05$).

we first compute the sample means and standard deviations.

$\bar{x}$ = mean of first sample

$= \frac{1}{7}(28+30+32+33+33+29+34)$

$= \frac{1}{7}(219) = \cancel{29.85}\ 31.286$

$\bar{y}$ = Mean of second sample

$= \frac{1}{6}(29+30+30+24+27+29)$

$= \frac{1}{6}(169) = 28.16$

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
28	-3.286	10.8	29	0.84	0.7056
30	-1.286	1.6538	30	1.84	3.3856
32	0.714	0.51	30	1.84	3.3856
33	1.714	2.94	24	-4.16	17.3056
33	1.714	2.94	27	-1.16	1.3456
29	-2.286	5.226	29	0.84	0.7056
34	2.714	7.366			
219		31.4358	169		26.8336

Now $s^2 = \frac{1}{n_1+n_2-2}\left[\sum(x_i-\bar{x})^2 + \sum(y_i-\bar{y})^2\right]$

$= \frac{1}{11}[31.4358 + 26.8336]$

$= \frac{1}{11}[58.2694]$

$= 5.23.$

$\therefore s = \sqrt{5.23}$

$= 2.3$

STUDENT'S t-TEST FOR DIFFERENCE OF MEANS :-

To test the significance difference between the sample means $\bar{x}$ and $\bar{y}$ of two independent samples of sizes n_1 and n_2 with the same variance, we use statistic.

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim n_1+n_2-2 \text{ d.f.}$$

where $\bar{x} = \frac{\Sigma x_i}{n}$, $\bar{y} = \frac{\Sigma y_i}{n}$

$$S^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x_i-\bar{x})^2+\Sigma(y_i-\bar{y})^2\right]$$

or

$$S^2 = \frac{1}{n_1+n_2-2}\left[(n_1-1)s_1^2+(n_2-1)s_2^2\right]$$

But for large samples, the following estimate is used.

$$S^2 = \frac{n_1 s_1^2+n_2 s_2^2}{n_1+n_2}$$

where s_1, s_2 are two sample standard deviations

① Two horses A and B were tested according to the time (in seconds) to run a particular track with the following results.

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29.	

Test whether the two horses have the same running capacity.

Sol:- Given $n_1 = 7, n_2 = 6$

1. Null Hypothesis $H_0: \mu_1 = \mu_2$

2. Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$

3. Level of significance, $\alpha = 0.05$

4. The test statistic is

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{31.286 - 28.16}{(2.3)\sqrt{\frac{1}{7} + \frac{1}{6}}} = 2.443.$$

Tabulated value of t for $7+6-2 = 11$ d.f at 5% l.o.s. is 2.2.

since cal. value t > Tab. value t, we reject the null hypothesis H_0.

and conclude that both horses A and B do not have the same running capacity.

(2) Find the maximum difference that we can expect with probability 0.95 between the means of samples of sizes 10 and 12 from a normal population if their S.D. are found to be 2 and 3 respectively.

Sol: we have $n_1 = 10$, $n_2 = 12$, $S_1 = 2$ & $S_2 = 3$

$$\therefore S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} = \frac{1}{10+12-2}\left[10(2)^2 + 12(3)^2\right]$$

$$= 7.4$$

$$\therefore S = \sqrt{7.4} = 2.72$$

1. Null Hypothesis $H_0: \mu_1 = \mu_2$

2. Alternative Hypothesis $H_1: \mu_1 \neq \mu_2$

3. level of significance $\alpha = 0.05$

Scanned with CamScanner

4. The test statistic is

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\Rightarrow |\bar{x} - \bar{y}| = |t| . s . \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= (2.086)(2.72)\sqrt{\frac{1}{10} + \frac{1}{12}}$$

$$= 2.43$$

[∵ Tabled value of t for $10+12-2 = 20$ d.f at 5% l.o.s is 2.086 (two-tailed)]

Hence the maximum difference between the means is 2.43.

3) To compare two kinds of bumper guards, 6 of each kind were mounted on a car and then the car was run into a concrete wall. The following are the costs of repair.

Guard 1	107	148	123	165	102	119
Guard 2	134	115	112	151	133	129

Use the 0.01 l.o.s. to test whether the difference between two sample means is significant.

4) The IQs (Intelligence quotients) of 16 students from one area of a city showed a mean of 107 with a S.D. of 10, while the IQs of 14 students from another area of the city showed a mean of 112 with a S.D. of 8. Is there a significant difference between the IQs of the two groups at a 0.05 l.o.s?

(2) Memory capacity of 10 students were tested before and after training state whether the training was effective or not from the following scores.

Before training	12	14	11	8	7	10	3	0	5	6
After training	15	16	10	7	5	12	10	2	3	8

(3) Scores obtained in a shooting competition by 10 soldiers before and after intensive training are given below:

Before	62	24	57	55	63	54	56	68	33	43
After	70	38	58	58	56	67	68	75	42	38

Test whether the intensive training is useful at 5% l.o.s.

SNEDECOR'S F-TEST OF SIGNIFICANCE

Test for Equality of Two population variances:-

Let two independent random samples of sizes n_1 and n_2 be drawn two normal populations.

To test the hypothesis that the two population variance σ_1^2 and σ_2^2 are equal.

The estimates of σ_1^2 and σ_2^2 are given by

$$S_1^2 = \frac{n_1 s_1^2}{n_1-1} = \frac{\sum(x_i-\bar{x})^2}{n_1-1} \text{ and } S_2^2 = \frac{n_1 s_2^2}{n_2-1} = \frac{\sum(y_i-\bar{y})^2}{n_2-1}$$

where s_1^2 and s_2^2 are the variances of the two samples.

Assuming that H_0 is true,

the test statistic $F = \frac{S_1^2}{S_2^2}$ or $\frac{S_2^2}{S_1^2}$ according as $S_1^2 > S_2^2$ or $S_2^2 > S_1^2$ follows F-distribution with (n_1-1, n_2-1) d.f.

① Pumpkins were grown under two experimental conditions. Two random samples of 11 and 9 pumpkins, show the sample standard deviations of their weights as 0.8 and 0.5 respectively. Assuming that the weight distributions are normal, test hypothesis that the true variances are equal.

Sol: Let the Null Hypothesis be

$H_0: \sigma_1^2 = \sigma_2^2$ i.e., the variances of sources are equal.

The Alternative Hypothesis is $H_1: \sigma_1^2 \neq \sigma_2^2$

Given $n_1 = 11$, $n_2 = 9$, $s_1 = 0.8$, $s_2 = 0.5$

Here samples S.D's are given

we have to determine the population variances S_1^2 and S_2^2

by using the relations.

$$n_1 s_1^2 = (n_1 - 1) S_1^2 \quad \text{and} \quad n_2 s_2^2 = (n_2 - 1) S_2^2$$

$$\Rightarrow S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{11 \times (0.8)^2}{10} = 0.704$$

$$\Rightarrow S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{9 \times (0.5)^2}{8} = 0.281$$

The test statistic is $F = \frac{S_1^2}{S_2^2} = \frac{0.704}{0.281}$

$= 2.5 \quad (\because S_1^2 > S_2^2)$

∴ calculated $F = 2.5$

Tab. value of F for (10,8) d.f at 5% l.o.S. is 3.35

Since calculated F < tabulated F, we accept the null hypothesis H_0 at 5% l.o.S. with (10,8) d.f and conclude that the variances of the two populations is the same and therefore, the two samples have the same variance.

PAIRED - SAMPLE t-TEST

If $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be the pairs of sales data before and after the sales promotion in a business concern. we apply paired t-test to examine the significance of the difference of the two situation.

Let $d_i = x_i - y_i$ or $y_i - x_i$ for $i = 1, 2, 3 \ldots n$

The test statistic for n paired observations (which are dependent) by taking the differences $d_1, d_2, \ldots d_n$ of the paired data.

$$t = \frac{\bar{d} - \mu}{S/\sqrt{n}} = \frac{\bar{d}}{S/\sqrt{n}} \quad (\because \mu = 0)$$

where $\bar{d} = \frac{1}{N} \Sigma d_i$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2$

$$\text{or } S^2 = \frac{\Sigma d^2 - n(\bar{d})^2}{n-1} \text{ or } \frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right]$$

are the mean and variance of the differences $d_1, d_2, \ldots d_n$ respectively and μ is the mean of the population of differences.

The above statistic follows student's t-distribution with $(n-1)$ d.f.

① The Blood Pressure of 5 women before and after intake of a certain drug are given below:

Before	110	120	128	132	125
After	120	118	125	136	121

Test whether there is significant change in Blood pressure at 1% l.o.s.

Sol: Let the Null hypothesis be $H_0: \mu_1 = \mu_2$
i.e., there is no significant difference in blood pressure.

before and after intake of drug.

The the Alternative Hypothesis $H_1: \mu_1 < \mu_2$

Assuming that Ho is true, the test statistic is

$$t = \frac{\bar{d}}{S/\sqrt{n}} \quad \text{where } \bar{d} = \frac{\Sigma d}{n}, \ d = y - x$$

$$\text{and } s^2 = \frac{\Sigma (d-\bar{d})^2}{n-1} = \frac{\Sigma d^2 - (\bar{d})^2 \times n}{n-1}$$

Calculations For $\bar{d}$ and S

womens	B.P. before intake of drug (x)	B.P. after intake of drug (y)	$d = y - x$	d^2
1	110	120	10	100
2	120	118	-2	4
3	123	125	2	4
4	132	136	4	16
5	125	121	-4	16
			$\Sigma d = 10$	$\Sigma d^2 = 140$

$$\therefore \bar{d} = \frac{\Sigma d}{n} = \frac{10}{5} = 2 \text{ and } s^2 = \frac{140 - (2)^2 \times 5}{4} = 30$$

$$\therefore S = \sqrt{30}$$

$$\therefore t = \frac{\bar{d}}{S/\sqrt{n}} = \frac{2}{\sqrt{30}/\sqrt{5}} = 0.82$$

Degree of freedom $= n-1 = 5-1 = 4$

Thus $t = 0.82 < 4.6$ at 1%. l.o.f with 4 d.f

Since the calculated value of t < the table value with 4 d.f. at 1% level, we accept Ho at 1% . l.o.s. and conclude that there is no significant change in Blood pressure after intake of a certain drug.

① Four methods are under development for making discs of a super conducting material. Fifty discs are made by each method and they are checked for super conductivity when cooled with liquid

	1st method	2nd method	3rd method	4th method
Superconductors	31	42	22	25
Failures	19	8	28	25

Test the significant difference between the proportions of super conductor at 5% level.

② From the following data, find whether there is any significant liking in the habit of taking soft drinks among the categories of employees.

Soft Drinks	Employees: Clerks	Teachers	Officers
pepsi	10	25	65
Thums up	15	30	65
Fanta	50	60	30

CHI-SQUARE TEST FOR POPULATION VARIANCE :-

Suppose that a random sample x_i $(i=1,2,\dots n)$ is drawn from a normal population with mean μ and variance σ^2. To test the hypothesis that the population variance σ^2 has a specified value σ_0

The test statistic is $\chi^2 = \sum \frac{(x_i-\bar{x})^2}{\sigma_0^2} = \frac{ns^2}{\sigma_0^2}$

where s^2 = sample variance $= \frac{\sum(x_i-\bar{x})^2}{n}$ and $ns^2 = (n-1)S^2$

Assuming that H_0 is true, the test statistic χ^2 follows χ^2-dist. with $(n-1)$ d.f.

Conclusion: If cal. value of $\chi^2 >$ tab value, we reject H_0 otherwise accept H_0.

① A firm manufacturing rivets wants to limit variations in their length as much as possible. The lengths (in cms) of 10 rivets manufactured by a new process are

2.15	1.99	2.05	2.12	2.17
2.01	1.98	2.03	2.25	1.93

Examine whether the new process can be considered superior to the old if the old population has S.D. 0.145 cm?

Sol: we have

$n = 10, \quad \bar{x} = \frac{\sum x_i}{n} = \frac{20.68}{10} = 2.068$

$s^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{0.09096}{10} = 0.0091$

and $\sigma_0 = 0.145$

1. Null Hypothesis $H_0 : \sigma^2 = \sigma_0^2$
2. Alternative " $H_1 : \sigma^2 > \sigma_0^2$
3. l.o.s. $\alpha = 0.05$
4. Assuming that H_0 is true, the test statistic is

$$\chi^2 = \frac{ns^2}{\sigma_0^2} = \frac{0.09096}{(0.145)^2} = 4.236 \qquad (\sigma_0 = 0.145)$$

$\therefore$ calculated $\chi^2 = 4.3$

d.f. $= n-1 = 10-1 = 9$

Tab. value χ^2 at 9 d.f at 5% l.o.s is 16.919

Since cal. value $\chi^2 <$ tab χ^2, we accept the null Hypothesis H_0

i.e., The new process cannot be considered superior to the old process.

② A random sample of size 20 from a normal population gives a mean of 42 and a variance of 25. Test the hypothesis that the population standard deviation is 8 at 5% level of significance.

PUPPalarajeshwari830@

$\therefore$ calculated $F = 4.075$

Tabulated value of F for $(n_2-1, n_1-1) = (5, 4)$ d.f at 5%. 1.0.5 is 6.26. since calculated F < tabulated F, we accept the null Hypoesis H_0 i.e., the variances are equal.

i.e., $\sigma_1^2 = \sigma_2^2$

③ Two random samples gave the following results

Sample	size	Sample mean	Sum of squares of Deviations from the mean
1	10	15	90
2	12	14	108

Test whether the samples came from the same normal population.

④ In one sample of 10 observations, the sum of the squares of the deviations of the sample values from sample mean was 120 and in the other sample of 12 observations, it was 314. Test whether. Test whether the difference is significant at 5%. level?

CHI-SQUARE (χ^2) TEST :-

Def:- If a set of events $A_1, A_2 \ldots A_n$ are observed to occur with frequencies $O_1, O_2 \ldots O_n$ respectively and according to probability rules $A_1, A_2 \ldots A_n$ are expected to occur with frequencies $E_1, E_2, \ldots E_n$ respectively with $O_1, O_2 \ldots O_n$ are called observed frequencies and $E_1, E_2 \ldots E_n$ are called expected frequencies.

If O_i $(i=1,2 \ldots n)$ is a set of observed (experimental) frequencies and E_i $(i=1,2 \ldots n)$ is the corresponding set of expected (theoretical) frequencies, then χ^2 is defined as

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \text{ with } (n-1) \text{ d.f.}$$

② The nicotine contents in milligrams in two samples of tobacco were found to be as follows:

Sample A	24	27	26	21	25	-
Sample B	27	30	28	31	22	36

Can it be said that the two samples have come from the same normal population.

Sol: Given $n_1 = 5$, $n_2 = 6$

Calculation for mean's and S.D's of the samples.

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
24	0.6	0.36	27	-2	4
27	2.4	5.76	30	1	1
26	1.4	1.96	28	-1	1
21	3.6	12.96	31	2	4
25	0.4	0.16	22	-7	49
			36	7	49
123		21.2	174		108

$$\therefore \bar{x} = \frac{\Sigma x}{n_1} = \frac{123}{5} = 24.6, \quad \bar{y} = \frac{\Sigma y}{n_2} = \frac{174}{6} = 29$$

$$\Sigma(x_i - \bar{x})^2 = 21.2 \qquad \Sigma(y_i - \bar{y})^2 = 108$$

$$\therefore S_1^2 = \frac{\Sigma(x_i-\bar{x})^2}{n_1 - 1} = \frac{21.2}{4} = 5.3 \text{ and}$$

$$S_2^2 = \frac{\Sigma(y_i - \bar{y})^2}{n_2 - 1} = \frac{108}{5} = 21.6$$

Null Hypothesis $H_0: \sigma_1^2 = \sigma_2^2$

The test statistic is $F = \frac{S_2^2}{S_1^2} = \frac{21.6}{5.3}$

$= 4.075$

Sol: Expected frequency of accidents each week $= \frac{100}{10} = 10$.

Null Hypothesis H_0: The accident conditions were the same during the 10 week period.

Alternative Hypothesis: The accidents conditions are different during the 10 week period.

Observed Frequency (O_i)	Expected Frequency (E_i)	$(O_i - E_i)$	$\frac{(O_i-E_i)^2}{E_i}$
12	10	2	0.4
8	10	-2	0.4
20	10	10	10.0
2	10	-8	6.4
14	10	4	1.6
10	10	0	0
15	10	-5	2.5
6	10	-4	1.6
9	10	-1	0.1
* 4	10	-6	3.6
100	10		
100	100		26.6

Now $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 26.6$.

i.e., calculated $\chi^2 = 26.6$

Here $n = 10$ observations are given

$\therefore$ D.f. $= n-1 = 10-1 = 9$

Tab. $\chi^2 = 16.9$

Since cal value $\chi^2 >$ Tab. χ^2. therefore, the Null Hypothesis is rejected and conclude that the accident conditions were not the same during the 10 week period.

Scanned with CamScanner

31

χ^2 is used to test whether differences between observed and expected frequencies are significant.

Chi-square Distribution is an important continuous prob. distribution and it is used both large and small tests. In chi-square tests, χ^2-distribution is mainly used.

(i) To test the goodness of fit

(ii) to test the independence of Attributes.

(iii) To test the population has a specified value of the variance σ^2.

① χ^2 TEST AS A GOODNESS OF FIT :-

Let $O_1, O_2 \dots O_n$ be a set of observed frequencies and $E_1, E_2 \dots E_n$ the corresponding set of expected frequencies.

Then the test statistic χ^2 is given by.

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i} \right] = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2} + \dots + \frac{(O_n-E_n)^2}{E_n}$$

Assuming Ho is true, the test statistic χ^2 follows chi-square distribution with $(n-1)$ d.f.

Conclusion: If the calculated value of χ^2 > tabulated value of χ^2 at α level, the Null Hypothesis Ho is rejected, otherwise Ho is accepted.

① The number of automobile accidents per week in a certain community are as follows: 12, 8, 20, 2, 14, 10, 15, 6, 9, 4. Are these frequencies in agreement with the belief that accident conditions were the same during this 10 week period.

① On the basis of information given below about the treatement of 200 patients suffering from a disease, State whethe the new treatement is comparatively superior to the conventi- -onal treatement.

	Favourable	Not favourable	Total
New	60	30	90
conventional	40	70	110

Sol:. Null Hypothesis H_0: No difference between new and conventional treatement (or) New and conventional trea- -tement are independent.

The number of degrees of freedom is $(2-1)(2-1)=1$

Expected frequencies are given in the table.

$$\text{Expected frequency} = \frac{\text{Row total} \times \text{column total}}{\text{Grand total}}$$

$\frac{90\times100}{200}=45$	$\frac{90\times100}{200}=45$	90
$\frac{100\times110}{200}=55$	$\frac{100\times110}{200}=55$	110
100	100	200

Caluclation of χ^2

observed frequen (O_i)	Expected frequen (E_i)	$(O_i-E_i)^2$	$(O_i-E_i)^2/E_i$
60	45	225	5
30	45	225	5
40	55	225	4.09
70	55	225	4.09
200	200		18.18

② A die is thrown 264 times with the following results. show that the die is biased. (Given $\chi^2_{0.05} = 11.07$ for 5 d.f.)

No. appeared on the die	1	2	3	4	5	6
Frequency	40	32	28	58	54	52

③ A pair of dice are thrown 360 times and the frequency of each sum is indicated below:

Sum :	2	3	4	5	6	7	8	9	10	11	12
Frequency :	8	24	35	37	44	65	51	42	26	14	14

would you say that the dice are fair on the basis of the chi-square test at 0.05 l.o.s.

② CHI-SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES:-

The test statistic

$$\chi^2 = \sum\left[\frac{(O_i - E_i)^2}{E_i}\right] \sim (r-1)(c-1) \text{ d.f.}$$

where expected frequency E_i of any cell

$$= \frac{\text{Row total} \times \text{column total}}{\text{Grand total.}}$$

(or)

The value of χ^2 is given by $\chi^2 = \dfrac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$

where ~~a+~~ $N = a+b+c+d$ with d.f. $= (2-1)(2-1) = 1$

$\frac{60\times40}{150}=16$	$\frac{30\times40}{150}=8$	$\frac{60\times40}{150}=16$	40
$\frac{60\times50}{150}=20$	$\frac{30\times50}{150}=10$	$\frac{60\times50}{150}=20$	50
$\frac{60\times60}{150}=24$	$\frac{30\times60}{150}=12$	$\frac{60\times60}{150}=24$	60
60	30	60	150

Calculation of χ^2

Observed frequency (O_i)	Expected Frequency (E_i)	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
15	16	1	0.0625
5	8	9	1.125
20	16	16	1
20	20	0	0
10	10	0	0
20	20	0	0
25	24	1	0.042
15	12	9	0.75
20	24	16	0.666
Total			3.6458

$$\therefore \chi^2 = \sum \frac{(O_i-E_i)^2}{E_i} = 3.6458$$

i.e., calculated $\chi^2 = 3.6458$

Table χ^2 for $(3-1)(3-1) = 4$ d.f at 5% l.o.s. is 9.488

Since calculated $\chi^2 <$ tabulated χ^2, we accept the null hypothesis H_0.

i.e., The hair colour and eye colour are independent

i.e., The hair colour and eye colour are not associated

$$\therefore \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 18.18$$

Tab value χ^2 for 1 d.f at 5% l.o.s is 3.841.

Since cal. χ^2 > tab χ^2 we reject the null Hypothesis Ho.

i.e., new and conventional treatement are not independent. The new treatement is comparatively superior to conventional treatement.

(2) The following table gives the classification of 100 workers according to sex and nature of work. Test whether the nature of work is independent of the sex of the worker.

	Stable	Unstable	Total
Males	40	20	60
Females	10	30	40
Total	50	50	100.

(3) Given the following contigency table for hair colour and eye colour. Find the value of χ^2. Is there good association between the two?

		Haircolour			
		Fair	Brown	Black	Total
Eyecolour	Blue	15	5	20	40
	Grey	20	10	20	50
	Brown	25	15	20	60
	Total	60	30	60	150

Sol: Null Hypothesis Ho: The two attributes, hair and eye colour are independent.

Table of expected frequencies:

$\frac{60\times40}{150}=16$	$\frac{30\times40}{150}=8$	$\frac{60\times40}{150}=16$	40
$\frac{60\times50}{150}=20$	$\frac{30\times50}{150}=10$	$\frac{60\times50}{150}=20$	50
$\frac{60\times60}{150}=24$	$\frac{30\times60}{150}=12$	$\frac{60\times60}{150}=24$	60
60	30	60	150

Calculation of χ^2

Observed frequency (O_i)	Expected Frequency (E_i)	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
15	16	1	0.0625
5	8	9	1.125
20	16	16	1
20	20	0	0
10	10	0	0
20	20	0	0
25	24	1	0.042
15	12	9	0.75
20	24	16	0.666
Total			3.6458

$$\therefore \chi^2 = \sum \frac{(O_i-E_i)^2}{E_i} = 3.6458$$

i.e., calculated $\chi^2 = 3.6458$

Table χ^2 for $(3-1)(3-1) = 4$ d.f at 5% l.o.s. is 9.488

Since calculated $\chi^2 <$ tabulated χ^2, we accept the null hypothesis H_0.

i.e., The hair colour and eye colour are independent

i.e., The hair colour and eye colour are not associated

Module

Co-relation and Regression

Co-relation

is a statistical analysis which measures and analysis the degree or extendt to 2 Variables flucuate with reference to each other.

Co-relation expresses the relationship or independence of set of Variables upon each other. one Variable may be called Subject (Independent) and the other relative (dependent).

Types of Co-relation:

(i) positive and negative
(ii) Simple and multiple
(iii) partial and total
(iv) linear and non-linear

TB

Methods of Studying Corelation

There are 2 different methods for finding out the relationship between Variables

(i) Graphical methods

(a) Scatter diagram or Scattergram.

(b) Simple Graph

(2) Mathematical methods

(a) karl's pearson's coefficient of corelation
(b) Spearmen's Rank Coefficient of Corelation
(c) Coefficient of Concurrent deviation.
(d) method of least Squares.

Scanned with CamScanner

Coefficient of Co-relation

is a Statistical technique used for analyzing the behaviour of 2 or more Variables. Its analyze deals with the association between 2 or more Variables.

Statistical measures of Co relation relates to Covariation between series but not of function or Casual relationship.

Karl Pearson's Coefficient of Co-relation:

Karl pearson — A British Biometrician and Statistican suggested a mathematical method for measuring the magnitude of linear relationship between 2 Variables. This is known as pearsonion Coefficient of Co relation and it is denoted by r.

There are Several formula to Calculate r.

(i) $$r = \frac{Cov(xy)}{\sigma_x \sigma_y}$$

(ii) $$r = \frac{\Sigma xy}{N\sigma_x \sigma_y}$$

(iii) $$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

(iv) $x = (x-\bar{x})$, $y = (y-\bar{y})$

where $\bar{x}, \bar{y}$ are mean's of the Series x and y

$\Sigma x \rightarrow \sigma$ of x, $\Sigma y \rightarrow \sigma$ of y

properties:

1) Coefficient of Co-relation lies between -1 and $+1$
Symbolically $-1 \leq r \leq 1$.

2) Coefficient of Co-relation is independent of change of Origin and Scale of measurements.

3) If X, Y are random Variables and a, b, c, d are any numbers such that $a \neq 0$, $c \neq 0$ then

$$r(ax+b, cy+d) = \frac{ac}{|ac|} r(x,y).$$

4) 2 independent Variables are un-corelated. i.e, if X and Y are independent Variables then $r(X,Y) = 0$.

problems:

1) Calculate the Coefficient of Corelation from following data

x	12	9	8	10	11	13	7
y	14	8	6	9	11	12	3

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

$X = x - \bar{x}$, $Y = y - \bar{y}$

$$\bar{X} = \frac{\Sigma x}{n} = \frac{70}{7} = 10$$

$$\bar{Y} = \frac{\Sigma y}{n} = \frac{63}{7} = 9$$

Scanned with CamScanner

X	Y	$x = X - \bar{X}$ $= X - 10$	$y = Y - \bar{Y}$ $= Y - 9$	x^2	y^2	xy
12	14	2	5	4	25	10
9	8	-1	-1	1	1	1
8	6	-2	-3	4	9	6
10	9	0	0	0	0	0
11	11	1	2	1	4	2
13	12	3	3	9	9	9
7	3	-3	-6	9	36	18
$\Sigma x = 70$	$\Sigma y = 63$			$\Sigma x^2 = 28$	$\Sigma y^2 = 84$	$\Sigma xy = 46$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$

$$= \frac{46}{\sqrt{28 \times 84}}$$

$r = 0.95$ [Highly co related] -

30.75

31

30.2

30

2) find if there is any Significant Co relation between the heights and weights given below

heights in inches	57	59	62	63	64	65	55	58	57
weights in lbS	113	117	126	126	130	129	111	116	112

Heights (H)	weights (w)	$H-\bar{H}$	$w-\bar{w}$	H^2	w^2	Hw
57	113	-3	-7	9	49	21
59	117	-1	-3	1	9	3
62	126	2	6	4	36	12
63	126	3	6	9	36	18
64	130	4	10	16	100	40
65	129	5	9	25	81	45
55	111	-5	-9	25	81	45
58	116	-2	-4	4	16	8
57	112	-3	-8	9	64	214
540	1080			102	472	216

Coefficient of Corelation $r = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$

$$\therefore r = \frac{216}{\sqrt{102 \times 471}}$$

$$= 0.98$$

3) Calculate the Coefficient of Correlation between age of Cars and the annual maintaince Cost and Comment

Age of Cars (year)	2	4	6	7	8	10	12
Annual maintaince (Rs)	1600	1500	1800	1900	1700	2100	2000

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \, \Sigma Y^2}}$$

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

$$Y = y - \bar{y}$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{49}{7} = 7$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{12600}{7} = 1800$$

x	y	$X = x - \bar{x}$	$Y = \frac{y-\bar{y}}{100}$	X^2	Y^2	XY
2	1600	-5	-2	25	4	100
4	1500	-3	-3	9	9	81
6	1800	-1	0	1	0	0
7	1900	0	1	0	1	0
8	1700	1	-1	1	1	1
10	2100	3	3	9	9	81
12	2000	5	2	25	4	100
49						

on Substitution we get

$$r = 0.83$$

Comment:

we observe that there is a high degree of Positive Co-relation between age of cars and annual maintaince Cost.

Rank Co-relation Coefficient

A British Psycologist Charles Edwards' pearmen ~~founded~~ out by the method of finding the Coefficient of Co relation by Ranks. This method is based on Rank and is useful for in dealing with qualitative characteristics such as morality, Character, intelligence and beauty. It Cannot be measured quantitatively as in the case of Pearson's Coefficient of Co-relation. It is based on the ranks given to the observations. It is applicable only to the individual Observations. The formula for Spearmen Rank Coeffici relation is given by

$$\rho = 1 - \frac{6 \Sigma d^2}{N(N^2-1)}$$

where ρ = Rank Coefficient of Co-relation.

d^2 = Sum of the Squares of the differences of 2 Ranks.

N = Number of paired observations.

properties:

1) The value of ρ lies between -1 and $+1$

2) If $\rho = 1$ there is complete Agreement in Order if the ranks and the directions of the rank is same.

3) If $\rho = -1$ then there is complete disagreement in the Order of the ranks and they are in opposite directio

problems:

1) The following are the rank obtained by 10 Studen in 2 Subjects, Statistics and mathematics. To what extend the knowledge of the Students in 2 Subjects is related?

Statistics	1	2	3	4	5	6	7	8	9	10
mathematics	2	4	1	5	3	9	7	10	6	8

sol:

Rank of Statistics (x)	Rank of maths (y)	$D = x - y$	D^2
1	2	-1	1
2	4	-2	4
3	1	2	4
4	5	-1	1
5	3	2	4
6	9	-3	9
7	7	0	0
8	10	-2	4
9	6	3	9
10	8	2	4
			$\Sigma d^2 = 40$

Scanned with CamScanner

$$\rho = 1 - \frac{6\Sigma d^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times \cancel{40}^{4}}{\cancel{10}(10^2-1)}$$

$$= 1 - \frac{24}{99}$$

$$= 0.75$$

2) The rank of 16 Students in Maths and Statistics are as follows:

(1,1) (2, 10) (3,3) (4,4) (5,5) (6,~~10~~) (7,2) (8,6) (9,8) (10,11) (11,15) (12,9) (13,14) (14,12) (15,16) (16,13).

Calculate the Rank-Co-relation Coefficient.

Solution Back

Sol:

Ranks in maths (x)	Ranks in Statistics (Y)	$D = X - Y$	D^2
1	1	0	0
2	10	-8	64
3	3	0	0
4	4	0	0
5	5	0	0
6	7	-1	1
7	2	5	25
8	6	2	4
9	8	1	1
10	11	-1	1
11	15	-4	16
12	9	3	9
13	14	-1	1
14	12	2	4
15	16	-1	1
16	13	3	9
		$\Sigma d = 0$	$\Sigma d^2 = 136$

Rank correlation coefficient

$$\rho = 1 - \frac{6 \Sigma d^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times 136}{16 \times 255}$$

$$= 1 - \frac{1}{5} = \frac{4}{5} = 0.8$$

Scanned with CamScanner

3) Random Sample of 5 College Students is Selected and their grades in maths and Statistics are found to be

	1	2	3	4	5
maths	85	60	73	40	90
statistics	93	75	65	50	80

Calculate pearson's Rank Corelation Coefficient.

marks in maths x	Rank x	marks in Stat y	Rank y	D = x-y	D^2
85	2	93	1	1	1
60	4	75	3	1	1
73	3	65	4	-1	1
40	5	50	5	0	0
90	1	80	2	-1	1
					$\Sigma d^2 = 4$

$$\rho = 1 - \frac{6 \Sigma d^2}{N(N^2-1)}$$

$$\rho = 1 - \frac{6 \times 4}{5(5^2-1)}$$

$$= 1 - \frac{1}{5}$$

$$= 0.8$$

Scanned with CamScanner

4) 10 Competitators in a musical test were ranked by the 3 judges a, b and c in the following order

***
**
v.Imp

Ranks By A	1	6	5	10	3	2	4	9	7	8
Ranks By B	3	5	8	4	7	10	2	1	6	9
Ranks by C	6	4	9	8	1	2	3	10	5	7

using Rank Co-relation method discuss which pair of judges has the nearest approach to common linkings in music.

Sol: Here $N = 10$

Ranks in A (X)	Ranks in B (Y)	Ranks in C(Z)	D_1 $x-y$	D_2 $x-z$	D_3 $y-z$	D_1^2	D_2^2	D_3^2
1	3	6	-2	-5	-3	4	25	9
6	5	4	1	2	1	1	4	1
5	8	9	-3	-4	-1	9	16	1
10	4	8	6	2	-4	36	4	16
3	7	1	-4	2	6	16	4	36
2	10	2	-8	0	8	64	0	64
4	2	3	2	1	-1	4	1	1
9	1	10	8	-1	-9	64	1	81
7	6	5	+1	2	1	1	4	1
8	9	7	-1	1	2	1	1	4

$\Sigma D_1 = 0$ $\Sigma D_2 = 0$ $\Sigma D_3 = 0$ $\Sigma D_1^2 = 200$,

$\Sigma D_2^2 = 60$, $\Sigma D^3 = 214$

$$\rho_1(X, Y) = 1 - \frac{6 \Sigma D_1^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times 200}{10 \times 99}$$

$$= 1 - \frac{40}{33}$$

$$= \frac{-7}{33}$$

$$\rho_2(X, Z) = 1 - \frac{6 \Sigma D_2^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times 60}{10 \times 99}$$

$$= 1 - \frac{4}{11}$$

$$= \frac{7}{11}$$

$$\rho_3(Y, Z) = 1 - \frac{6 \Sigma D_3^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times 214}{10 \times 99}$$

$$= 1 - \frac{214}{165}$$

$$= -\frac{49}{165}$$

Since $\rho_2(X, Z)$ is maximum, we conclude that the pair of judges A and C has the nearest approach to common likings to music.

Equal and repeated Ranks :

If any 2 or more persons are braketed equal in any classification or if there is more than 1 item with the Same value in the Series then the Spearmen formula for calculating the Rank-Co relation coefficient breaks down. In this case Common ranks are given to repeated items. The common rank is the average of the ranks if these items would have assumed, if they were difficult from each other and the next item will get rank next to ranks already assumed.

$$\rho = 1 - 6\left[\frac{\Sigma d^2 + \frac{1}{12} m(m^2-1) + \frac{1}{12} m(m^2-1) + \dots}{N(N^2-1)}\right]$$

1) from the following data calculate the Rank Co-relate Coefficient after making adjustment for tied ranks.

x	48	33	40	9	16	16	65	24	16	57
y	13	13	24	6	15	4	20	9	6	19

X	Rank (x)	Y	Rank (y)	D = x−y	D^2
48	3	13	5.5	−2.5	6.25
33	5	13	5.5	−0.5	0.25
40	4	24	1	3	9
9	10	6	8.5	1.5	2.25
16	8	15	4	4	16
16	8	4	10	−2	4
65	1	20	2	−1	1
24	6	9	7	−1	1
16	8	6	8.5	−0.5	0.25
57	2	19	3	−1	1
					$\Sigma D^2 = 41$

16 is repeated 3 times in x items

Hence $m = 3$.

Since 13 and 6 are repeated twice in y items

Hence $m = 2$, $m = 2$

$$\therefore \rho = 1 - \frac{6\left[\Sigma d^2 + \frac{1}{12}(m^3-1) + \frac{1}{12}(m^3-1) + \frac{1}{12}(m^3-1)\right]}{N(N^2-1)}$$

$$= 1 - \frac{6\left[41 + \frac{1}{12}\left[3^3-3\right) + 1/12\left(2^3-2\right) + 1/12\left(2^3-2\right)\right]}{10(10^2-1)}$$

$$= 0.7$$

Regression Equation

→ It's an algebric expression of the regression line. It can be classified into Regression equation, Regression Co-efficient, Individual observation and Group discussion.

The Standard form of regression equation is

$Y = a + bX$ $\quad$ $\underline{y = mx + c}$

where a, b are called Constants.

a indicates the Value of y when $x = 0$. It is called y-intercept

b indicates the Value of Slope of the regression line and gives a measure of change of Y for a (unique) change in X. It is also called the Regression Coefficient of Y on X. The Values of a and b found with the help of the following normal equations.

Regression equation of Y on X

$$Y = a + bX$$

$$\Sigma Y = na + b\Sigma X$$

$$\Sigma XY = a\Sigma X + b\Sigma X^2$$

Regression equation of X on Y

$$X = a + bY$$

$$\Sigma X = na + b\Sigma Y$$

$$\Sigma XY = a\Sigma Y + b\Sigma Y^2$$

Deviation taken from Arithmetic mean of X on Y

1) Regression equation of X on Y

$$X - \bar{X} = r.\frac{\sigma_x}{\sigma_y}(Y - \bar{Y})$$

where $\bar{X}$ = mean of X Series

$\bar{Y}$ = mean of Y Series

The regression Coefficient of X on Y $= r.\frac{\sigma_x}{\sigma_y}$

$$r.\frac{\sigma_x}{\sigma_y} = \frac{\Sigma xy}{\Sigma \cancel{x} y^2} = bxy$$

now formula becomes

$$X - \bar{X} = \cancel{x}.\frac{\Sigma xy}{\Sigma \cancel{x} y^2}(Y - \bar{Y})$$

The Regression Coefficient of Y on X is

$$Y - \bar{Y} = r.\frac{\sigma_y}{\sigma_x}(X - \bar{X})$$

The regression Coefficient y on X

$$= r.\frac{\sigma_y}{\sigma_x} = \frac{\Sigma xy}{\Sigma x^2} = byx$$

Thus $r^2 = bxy\, byx$

problems:

1) A panel of 2 judges P and Q graded 7 dramatic performances by independently awarding marks as follows:

performance	1	2	3	4	5	6	7
Marks by P	46	42	44	40	43	41	45
Marks by Q	40	38	36	35	39	37	41

The 8^{th} performance, which judge Q would not attend, was awarded 37 marks by judge P. If judge Q had also being present how many marks would be expected to have being awarded by him to the 8^{th} Performance.

Sol:- Now we have to find out regression equation of y on x.

$$y - \bar{y} = byx\,(x - \bar{x}) \quad \text{--- (1)}$$

$$\text{where } byx = \frac{\Sigma xy}{\Sigma x^2}$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{301}{7} = 43$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{266}{7} = 38$$

$$byx = \frac{21}{28} = 0.75$$

performance	marks by P x	$X = x-\bar{x}$	Marks by Q y	$Y = y-\bar{y}$	XY	X^2
1	46	3	40	2	6	9
2	42	-1	38	0	0	1
3	44	1	36	-2	-2	1
4	40	-3	35	-3	9	9
5	43	0	39	1	0	0
6	41	-2	37	-1	2	4
7	45	2	41	3	6	4
	$\Sigma x = 301$		$\Sigma y = 266$		$\Sigma XY = 21$	$\Sigma X^2 = 28$

// we know $Y - \bar{Y}^2 = bYX\,(X - \bar{X})$ —— ① [known]

Sub. the Value in ①

$$Y - 38 = 0.75\,(X - 43)$$

$$Y - 38 = 0.75X - 32.25$$

$$Y = 0.75X - 32.25 + 38$$

$$Y = 0.75X + 5.75$$

If $X = 37$

then $Y = 0.75\,(37) + 5.75$

$$= 33.5$$

Hence Q would have been present, he would have awarded 33.5 marks to the 8th performance.

2) find the most likely production to a rainfall 40

from the following data

	Rainfall (x)	production (Y)
Average	30	500 Kgs
σ	5	100 Kgs
Coefficient of Co relation	0.8	

Y on x

Given $\bar{x} = 30$, $\bar{Y} = 500$

$\sigma_x = 5 \quad \sigma_y = 100$

$r = 0.8$

now we have to find y when x=40

The regression line y on x is

$$Y - \bar{Y} = r . \frac{\sigma_x}{\sigma_y} (x - \bar{x})$$

$$\Rightarrow \cancel{Y - x\ 30 = 0.8} \left(\frac{\cancel{5}\ 5}{100}\right)(\cancel{x-30})$$

by Sub... the values we get

$$Y - 500 = 0.8 \left(\frac{5}{100}\right) (x - 30)$$

but also given x = 40

$$Y - 500 = 0.8 \left(\frac{5}{100}\right) (40 - 30)$$

$$Y = 500.4$$

HW

Determine the equation of straight line which best fits the data

x	10	12	13	16	17	20	25
y	10	22	24	27	29	33	37

4) find ~~the~~ regression lines from the following data

x	10	12	13	12	16	15
y	40	38	43	45	37	43

Solutions:-

3) we know the Standard form of Regression equation is $Y = a + bX$

X	Y	XY	X^2
x	y	xy	x^2
10	10	100	100
12	22	144	264
13	24	169	312
16	27	256	432
17	29	289	493
20	33	400	660
25	37	625	925

$\Sigma y = 182$ $\quad\quad \Sigma xy = 3186$

we know $\Sigma x = 113$ and $\Sigma x^2 = 1938$ [by calculating]

4)

$\cancel{x}$	y	X ($x-\bar{x}$)	Y ($y-\bar{y}$)	XY	X^2	Y^2
10	40	-3	-1	3	9	1
12	38	-1	-3	3	1	9
13	43	0	2	0	0	4
12	45	-1	4	-4	1	16
16	37	3	-4	-12	9	16
15	43	2	2	4	4	4
$\Sigma x = 78$	$\Sigma y = 246$			-6	24	50

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$

$$\bar{X} = \frac{\Sigma x}{n} = \frac{78}{6} = 13$$

$$\bar{Y} = \frac{\Sigma y}{n} = \frac{246}{6} = 41$$

$$X - \bar{X} = \frac{\Sigma XY}{\Sigma Y^2}(Y - \bar{Y})$$

$$X - 13 = \frac{-6}{50}(Y - 41)$$

Regression equation of Y on X is $X - 13 =$

$$Y - \bar{Y} = r\frac{\sigma_y}{\sigma_x}(X - \bar{X}) \quad \text{——} \; ①$$

$\cancel{X =}$

now

$$r\frac{\sigma_y}{\sigma_x} = \frac{\Sigma XY}{\Sigma X^2} = -0.25$$

$$Y - 41 = -0.25(X - 13) \; [\text{by } 1]$$

$$\Rightarrow Y = -0.25X + 44.25$$

when X is 20 and Y = 39.25

when the price is Rs 20, the likely demand is 39.25

$\Sigma Y = b\Sigma X + NA$

$\Sigma Y = 182, \Sigma X = 113; N=7$

$113b + 7a = 182$ — (1)

$\Sigma XY = 3186, \Sigma X^2 = 1983; \Sigma X = 113$

$1983b + 113a = 3186$ — (2)

multiplying (1) by 113,

$12769b + 791a = 20566$ — (3)

multiplying (2) by 7

$13881b + 791a = 22302$ — (4)

$(4) - (3) \Rightarrow b = \frac{1736}{1112} = 1.56$

$a = 0.82$

∴ The equation of Straight line is

$Y = a + bX$

$a = 0.82; b = 1.56$

$Y = 0.82 + 1.56X$

∴ The equation of the required Straight line is $Y = 0.82 + 1.56X$

This is called regression equation of Y on X

Scanned with CamScanner

5M Angle Between 2 Regression lines

Let the Regression line of X on Y is $x-\bar{x} = r.\frac{\sigma_x}{\sigma_y}(y-\bar{y})$

The Slope $m_1 = \frac{1}{r}\frac{\sigma_y}{\sigma_x}$

$\boxed{\text{IB proof wrong}}$

The Regression line of Y on X is $y-\bar{y} = r.\frac{\sigma_y}{\sigma_x}(x-\bar{x})$

Slope $m_2 = r.\frac{\sigma_y}{\sigma_x}$

$y = m_1x + C$

$y = m_2x + C$

$$\tan\theta = \frac{m_1 - m_2}{1 + m_1 m_2}$$

$$\tan\theta = \frac{m_1 - m_2}{1 + m_1 m_2}$$

$$= \frac{\frac{1}{r}.\frac{\sigma_y}{\sigma_x} - r\frac{\sigma_y}{\sigma_x}}{1 + \frac{1}{\cancel{r}}\frac{\sigma_y}{\sigma_x} \times \cancel{r}\frac{\sigma_y}{\sigma_x}}$$

$$= \frac{\frac{\sigma_y}{\sigma_x}\left(\frac{1-r^2}{r}\right)}{1 + \frac{\sigma_y^2}{\sigma_x^2}}$$

$$= \frac{\frac{\sigma_y}{\cancel{\sigma_x}}\left(\frac{1-r^2}{r}\right)}{\frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^{\cancel{2}}}} = \frac{\sigma_y\sigma_x\left(\frac{1-r^2}{r}\right)}{\sigma_x^2 + \sigma_y^2}$$

problem:

1) If θ is the angle between 2 regression lines and σ of y is twice the σ of x and $r = 0.25$ find $\tan\theta$ $\quad \sigma_y \quad \sigma_x$

sol: Given $\sigma_y = 2\sigma_x$, $r = 0.25$

$$\tan\theta = \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\left(\frac{1-r^2}{r}\right)$$

$$= \frac{\sigma_x \cdot 2\sigma_x}{\sigma_x^2 + 4\sigma_x^2}\left(\frac{1-(0.25)^2}{0.25}\right)$$

$$= \frac{2\sigma_x^2}{5\sigma_x^2}\left(\frac{1-(0.25)^2}{0.25}\right)$$

$$\tan\theta = \frac{2}{5}\left(\frac{1-(0.25)^2}{0.25}\right) \qquad \theta = \tan^{-1}(1.5)$$

$$\tan\theta = 1.5$$

2) If σ of x is σ of y is Σ and the angle between regression lines is 4/3. find r

Given $\sigma_x = \sigma_y = \sigma$

$$\tan\theta = \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\left(\frac{1-r^2}{r}\right)$$

$$\tan\theta = \frac{\sigma^2}{\sigma^2 + \sigma^2}\left(\frac{1-r^2}{r}\right)$$

$$\theta = \tan^{-1}\left(\frac{\cancel{\sigma^2}}{2\cancel{\sigma^2}}\left(\frac{1-r^2}{r}\right)\right)$$

$$\theta = \tan^{-1}\left(\frac{1}{2}\left(\frac{1-r^2}{r}\right)\right)$$

By data $\theta = \tan^{-1}(4/3)$

$$\frac{4}{3} = \frac{1}{2}\left(\frac{1-r^2}{r}\right)$$

$$\frac{1-r^2}{r} = \frac{8}{3}$$

$$3-3r^2 = 8r$$

$$3r^2+8r-3=0$$

$$3r^2+9r-r-3=0$$

$$3r(r+3)-1(r+3)=0$$

$$(3r-1)(r+3)=0$$

$$r = 1/3,\ r=-3$$

$$\therefore r = 1/3 \quad (r \neq -3)$$

$$[\because -1 \leq r \leq 1]$$

3) Find the mean Values of the Variable x and y and the Co-relation coefficient from the following regression equations.

$$2y - x - 50 = 0$$

$$3y - 2x - 10 = 0$$

Given

$$2y - x - 50 = 0 \quad \text{——— (1)}$$

$$3y - 2x - 10 = 0 \quad \text{——— (2)}$$

multiply 2 by (1)

$$4y - 2x - 100 = 0$$

$$\underline{-\; 3y \ominus 2x \ominus 10 = 0}$$
$$\quad\;\; + \quad\;\; +$$

$$y - 90 = 0$$

$$y = 90$$

Sub y in (1)

$$2(90) - x - 50 = 0$$

$$x = 130$$

$$\therefore \bar{x} = 130,\ \bar{y} = 90$$

Rewriting (1) and (2)

$$y = \frac{1}{2}x + 25 \ \& \ x = \frac{3}{2}y - 5$$

$$r\frac{\sigma_y}{\sigma_x} = \frac{1}{2} \quad \text{and} \quad r \cdot \frac{\sigma_x}{\sigma_y} = \frac{3}{2}$$

$$\therefore r^2 \frac{\sigma_y}{\sigma_x} \cdot \frac{\sigma_x}{\sigma_y} = \frac{1}{2} \cdot \frac{3}{2}$$

$$r^2 = 3/4 \qquad r = \sqrt{\frac{3}{4}} = 0.866 \text{ (Ans)}$$

$$2Y = X + 50$$

$$Y = \frac{1}{2}X + 25$$

$$2X = 3Y - 10$$

$$X = 3/2\, Y - \frac{10}{\cancel{2}}$$

$$r \cancel{\frac{\sigma_x}{\sigma_x}}\, r \cdot \cancel{\frac{\sigma_x}{\sigma_y}} = 1/2 \cdot 3/2$$

$$r^2 = 3/4$$

$$r = \sqrt{3}/2$$

$$= 0.866$$

1/2/17 (i) If $X = 2y + 3$ and $Y = kx + 6$ are the regression lines of X on Y and Y on X respectively.

(a) Show that $0 \le k \le 1/2$

(b) If $k = \frac{1}{8}$ find r and $\bar{x}, \bar{y}$

Sol:- Given $X = 2Y + 3$ is the regression line of X on Y

$$\therefore \quad r\frac{\sigma_x}{\sigma_y} = 2$$

The regression line on Y on X is

$$Y = kx + 6$$

$$\therefore \quad r\frac{\sigma_x}{\sigma_y} = k$$

multiply this two

$$r.\frac{\sigma_x}{\sigma_y} \cdot r\frac{\sigma_y}{\sigma_x} = 2k$$

$$r^2 = 2k$$

we have $0 \le r^2 \le 1$

$\Rightarrow \; 0 \le 2k \le 1$ [divide 1/2 on both the sides]

$\Rightarrow 0 \le k \le \frac{1}{2}$

(b) $k = \frac{1}{8}$ then $Y = \frac{1}{8}x + 6$

$$r^2 = 2k$$

$$r^2 = \cancel{2}\left(\frac{1}{\cancel{8}4}\right)$$

$$r = \pm\frac{1}{2}$$

Since both regression Coefficients are +ve we take +ve Values of r

$\therefore r = \frac{1}{2}$ (+ve)

by Substitution.

$\therefore \bar{X} = 20$ and $\bar{Y} = 8.5$