Unit I

Random Variables and Probability Distributions

Introduction to probability:

The term 'probability' was named by Pierre Simon laplace in 1795. J. Cardan an Italian physician & mathematican wrote the first book on probabilily named the book of games chance.

Experiment ~~/trial~~ : It is defined as act of doing something.

Eg: throwing a coin; throwing a die

Random Experiment:

we have two types of experiments

(i) deterministic or perdictable

(ii) undeterministic or unpredictable (random experiment)

Def'n: An experiment is called deterministic experiment if, the result can be predicted with certain prior to the performance of the experiment.

Example: Throwing a stone upwards where it is known that the stone will definitely fall due to gravitational force.

Def'n: An experiment is called random expt, if when repeated under the same conditions, it is such that the outcome cannot be predicted with certainity; but all possible outcomes can be determined prior to the performance of the expt.

Note: Here experiment means a random experiment.

Trial: A single performance of an experiment is called a trial.

Event | outcome: It is defined as the result of the experiment.

Eg: In throwing a die, getting (1 or 2 or 3 .. or 6) is an event.

In tossing a coin getting Head or tail is an event.

Event may be classified as two types:

Elementary event: It is an event which cannot be broken further into smaller events.

Eg: In throwing a die, each of the event is an elementary event.

Compound event: It is an event which can be broken further into smaller events.

Eg: In throwing a die, the event that an odd number turns up is a combination of three smaller events. 1 or 3 or 5 turns up.

Sample Space: The collection of all possible outcomes of an random experiment is called the ~~sp.~~ sample space, denoted by 'S'.

Eg: In a random expt of drawing one card from a pack of 52 cards, outcome is any particular card. Hence, the sample space S consists of all individual cards and $n(S) = 52$

Exhaustive Events: The total number of possible outcomes of a random experiment.

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

Favourable events: The events which are favourable to a particular event of an expt are called favourable events.

Eg: when a die is rolled getting an even number.

Note: Events which are favourable to a particular event of an expt are called successes and the remaining are called as failures.

Mutually Exclusive / Disjoint events: If no two or more of the events can happen simultaneously in the same trial.

Independent events: Two events are said to be independent if the occurrence (Non-occurrence) of one event has no influence on the occurrence (Non-occurrence) of the other event.

Eg: A card with replacement.

If we draw a card from a pack of well shuffled cards and replace it before drawing the second card, the result of the second draw does not affect the result of the first draw.

Dependent events: Two events are said to be dependent if the occurrence (Non-occurrence) of one event has influence on the occurrence of the other event.

Eg: A card without replacement

If we draw a card from a pack of well shuffled cards and do not replace it before the second ~~dr~~ card is drawn. Then the result of the second draw is affected by the result of the first draw.

Classical def'n of probability:

If there are 'n' mutually exclusive and equally likely events of a random experiment, out of which, 'm' events are favourable for a particular event E, then we define the probability of E, as

$$P(E) = \frac{m}{n} = \frac{\text{Number of favourable events w.r.t to } E}{\text{Number of total events of the expt.}}$$

This prob. is also known as prob. of success of E.

In this expt, 'm' results are favourable to E, and hence the remaining $n-m$ results are not favourable to the event E. This set of unfavourable events denoted by $\bar{E}$.

$\therefore$ Prob. of $P(\bar{E}) = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - P(E)$

$\therefore \quad P(E) + P(\bar{E}) = 1$

Note: i) Generally prob. of success & prob. of failure are denoted by p & q respectively.

$\therefore \quad p + q = 1$

ii) If $P(E) = 1$, E is called a certain or sure event and if $P(E) = 0$, E is called an impossible event.

Axioms of probability:

If S is the sample space, and E is any event in a random experiment,

i) $0 \leq P(E) \leq 1$ for each event E in S

ii) $P(S) = 1$

iii) If E_1 & E_2 any mutually exclusive, ents in S, then

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

Some elementary theorems:

Addition theorem of probability for two independent events:

If A & B are two independent events then $P(A \cup B) = P(A) + P(B)$

Addition theorem of probability for two dependent events:

If A & B are two dependent events then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Conditional probability: If A and B are two events of a sample space S, then the event B can only happen if and only if the event A is already happenend.

It is usually denoted by $P(B|A)$

(or) $P(A|B)$ which means event A can only happen iff if the event B is already happened.

Multiplication theorem of probability / Joint Probability
(For two independent events)

If A & B are independent events, then
$P(A \text{ and } B) = P(A \cap B) = P(A) \cdot P(B)$

Joint probability for two dependent events:

If A and B are two dependents events,
then $P(A \text{ and } B) = P(A \cap B) = P(A) \cdot P(B/A)$

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

Problems:

1. A man's pocket contains 5 fifty paise, 4 twenty paise & 4 ten paise coins. A boy is asked to draw two coins at random. What is the prob. of the boy drawing

(a) Max. possible amount

(b) Min. possible amount

(c) coins of different value

Sol There are total of 13 coins in the pocket and two coins are drawn at random.

∴ Total no: of ways 13 coins are drawn are = $^{13}C_2$ ways

(i) let E be the event of drawing max. amt

ie, $m = {}^{5}C_2$

$$\therefore P(E) = \frac{m}{n} = \frac{^{5}C_2}{^{13}C_2}$$

(ii) let E be the event of drawing min. amt

ie, $m = {}^{4}C_2$

$$\therefore P(E) = \frac{m}{n} = \frac{^{4}C_2}{^{13}C_2}$$

(iii) let E be the event of drawing coins of different value.

ie, $m = {}^{5}C_1 \times {}^{4}C_1 + {}^{4}C_1 \times {}^{4}C_1 + {}^{4}C_1 \times {}^{5}C_1$

$$\therefore P(E) = \frac{m}{n} = \frac{^{5}C_1 \times {}^{4}C_1 + {}^{4}C_1 \times {}^{4}C_1 + {}^{4}C_1 \times {}^{5}C_1}{^{13}C_2}$$

→ 2. Three electronic lamps are fitted in a room. Three bulbs are selected random from 10 bulbs having 6 good bulbs. what is the chance that the room is lighted?

Sol ~~Let~~ Total bulbs are 10; of which 3 are selected at random.

i.e, Total possible outcomes are

$$n = {}^{10}C_3$$

Given out of 10 bulbs, 6 bulbs are good.

A.L P(Good bulb) = $\frac{6}{10}$

Let E be the event that the room is lighted. i.e, the bulb is good.

$$\therefore m = {}^6C_1 \times {}^4C_2 + {}^6C_2 \times {}^4C_1 + {}^6C_3$$

$$\therefore P(E) = \frac{{}^6C_1 \times {}^4C_2 + {}^6C_2 \times {}^4C_1 + {}^6C_3}{{}^{10}C_3}$$

3. In a group there are 3M and 2W. Three persons are selected at random from this group. Find the prob. that one man and two women or two men and one women are selected.

Sol Total persons in a group are = 5

Three people are selected at random

${}^5C_3$

let E be the event of selecting
(1m & 2w) or (2m & 1w)

$m = {}^{3}C_{1} \times {}^{2}C_{2} + {}^{3}C_{2} \times {}^{2}C_{1}$

$$P(E) = \frac{m}{n} = \frac{{}^{3}C_{1} \times {}^{2}C_{2} + {}^{3}C_{2} \times {}^{2}C_{1}}{{}^{5}C_{2}}$$

3. If 3 of 20 tires in a storage are defective and 4 of them are randomly selected for inspection, what is the prob. that only one of the defective tires will be included?

sol There are ${}^{20}C_{4}$ ways of selecting 4 tires out of 20.

Then let E be the event that only one of the defective tires will be included.

$m = {}^{3}C_{1} \times {}^{17}C_{3}$

$$P(E) = \frac{{}^{3}C_{1} \times {}^{17}C_{3}}{{}^{20}C_{4}}$$

~~4. Twelve balls are distributed at random among three boxes. what is the prob that the first box will contain 3 balls?~~

~~Sol~~

5. Out of 15 items 4 are not in good condition 4 are selected at random. Find the prob that i) All are not good
ii) Two are not good.

Sol Total no: of items = 15

4 are selected at random = $^{15}C_4$ ways

i) let E be the event of drawing not good items,

ie, $m = {}^{4}C_4$

$$\therefore P(E) = \frac{^{4}C_4}{^{15}C_4}$$

ii) let E be the event of selecting two items which are not good.

ie, $m = {}^{11}C_2 \times {}^{4}C_2$

$$\therefore P(E) = \frac{^{11}C_2 \times {}^{4}C_2}{^{15}C_4}$$

6. A bag contains 4 G, 6 B, 7 W balls. A ball is drawn at random. what is the prob. that it is either a green or a black ball.

Sol Total balls are 17, of which one ball is drawn at random.

$\therefore\ n = 17_{C_1}$

Let E_1 be the prob of drawing a green ball ie, 4_{C_1}

$P(E_1) = \frac{4}{17}$

let E_2 be the prob. of drawing a black ball ie, 6_{C_1}

ie, $P(E_2) = \frac{6}{17}$

$\therefore$ Prob. of drawing either a green or black ball is $P(E_1 \cup E_2) = P(E_1) + P(E_2)$

$= \frac{4}{17} + \frac{6}{17} = \frac{10}{17}$

7. A bag contains 8R and 6B balls. Two drawings of each 2 balls are made. Find the prob. that the first drawing gives two R balls and second drawing gives 2 B balls, if the balls drawn first are replaced before the second draw.

Sol let A_t be the event of drawing 2 R balls in the first draw from the bag of 8R & 6B

$$P(A) = \frac{^8C_2}{^{14}C_2}$$

If two balls are replaced.

let B be the event of drawing 2 B balls in the second draw.

$$P(B) = \frac{^6C_2}{^{14}C_2}$$

$\therefore$ Prob. of drawing 2R in first draw & 2B in second draw

i.e, $P(A \cap B) = P(A) \cdot P(B)$

$$= \frac{^8C_2}{^{14}C_2} \cdot \frac{^6C_2}{^{14}C_2}$$

→ Twelve balls are distributed at random among three boxes. What is the prob. that first box will contain 3 balls?

Sol Since each ball can go to any one of the 3 boxes there are 3 ways in which a ball can go to any one of the three boxes.

∴ There are 12 balls to be distributed in 3 boxes

Hence 3^{12} ways, it can be replaced.

Number of ways in which 3 balls out of 12 can go in first box $^{12}C_3$

∴ Remaining 9 balls are placed in 2^{nd} & 3^{rd} boxes in 2^9 ways.

∴ Total number of ways $n = 3^{12}$

Favourable cases: $^{12}C_3 \times 2^9$.

$$\frac{^{12}C_3 \times 2^9}{3^{12}}.$$

→ A bag contains 6W and 9B balls. Four balls are drawn at a time. Find the prob. for first draw to give 4W and the second to give 4B balls, when the balls are not replaced before the second draw.

Sol Total balls 15 of which 4 balls are drawn.

$n = {}^{15}C_4$

let A be the event of drawing 4W balls in the first draw.

$m = {}^{6}C_4$

$$\therefore \quad P(A) = \frac{^{6}C_4}{^{15}C_4}$$

Given the balls are not replaced before 2^{nd} draw. (Remaining 11 balls)

$\therefore \quad n = {}^{11}C_4$

let B be the event of drawing 4 black ball in the 2^{nd} draw.

$m = {}^{9}C_4$

$$P(B/A) = \frac{{}^{9}C_4}{{}^{11}C_4}$$

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

$$= \frac{{}^{6}C_4}{{}^{15}C_4} \cdot \frac{{}^{9}C_4}{{}^{11}C_4}$$

Baye's theorem: If $E_1, E_2, \ldots E_n$ are mutually exclusive events with $P(E_i) \neq 0$ for each i and A is an arbitrary event intersecting with every E_i such that $P(A) > 0$ then

$$P(E_1|A) = \frac{P(E_1) \cdot P(A|E_1)}{P(E_1) \cdot P(A|E_1) + P(E_2) \cdot P(A|E_2) + \ldots P(E_n)\, P(A|E_n)}$$

$$\vdots$$

$$P(E_i|A) = \frac{P(E_i) \cdot P(A|E_i)}{\sum_{i=1}^{n} P(E_i)\, P(A|E_i)}$$

(or) If E_1, E_2 are two mutually exclusive and collectively exhaustive events and B is a simple event which intersects E_1, E_2 $\ni$ $P(B) \neq 0$ then

$$P(E_1|B) = \frac{P(E_1) \cdot P(B|E_1)}{P(E_1) \cdot P(B|E_1) + P(E_2) \cdot P(B|E_2)}$$

$$P(E_2|B) = \frac{P(E_2) \cdot P(B|E_2)}{P(E_1) \cdot P(B|E_1) + P(E_2) \cdot P(B|E_2)}$$

Note: i) $P(E_1), P(E_2) \ldots P(E_n)$ are called "<u>apriori</u>" prob's

ii) $P(A|E_1), P(A|E_2) \ldots P(A|E_n)$ are called "<u>likelihood</u>" prob's

iii) $P(E_1|A), P(E_2|A) \ldots P(E_n|A)$ are called "<u>posteriori</u>" prob's

'' bag A contains 2W and 3R balls and a bag 'B' contains 4W and 5R balls. One ball is drawn at random from one of the bags and it is found to be red. Find the prob. that the red ball is from bag B.

Sol let A & B denote the events of selecting bag A & B respectively.

$$P(A) = \frac{1}{2} \ ; \ P(B) = \frac{1}{2}$$

let E denote the event of drawing red ball Having selected bag A, the prob. of drawing a red ball from A $= P(E|A) = \frac{3}{5}$

$|||^{ly}$ $P(E|B) = \frac{5}{9}$

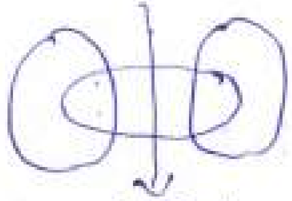
one of the bags is selected at random and from it a ball is drawn at random which is found to be red.

$$P(B|E) = \frac{P(B)\cdot P(E|B)}{P(A)\cdot P(E|A) + P(B)\cdot P(E|B)}$$

$$= \frac{\frac{1}{2}\cdot\frac{5}{9}}{\frac{1}{2}\cdot\frac{3}{5} + \frac{1}{2}\cdot\frac{5}{9}} = \frac{25}{52}$$

→ Suppose 5M out of 100 and 25W out of 10,000 are colour blind. A colour blind person is chosen at random. What is the prob. of the person being a male.

Sol let A, B be the events of Male and female. Then



$P(A) = P(B) = \frac{1}{2}$

let E be the event of blind person.

$P(E|A) = \frac{5}{100} = 0.05$; $P(E|B) = \frac{25}{10,000} = 0.0025$

The prob. that the chosen person is male given that he is blind.

i.e., $P(A|E) = \dfrac{P(A) \cdot P(E|A)}{P(A) \cdot P(E|A) + P(B) \cdot P(E|B)}$

$= \dfrac{\frac{1}{2} \cdot 0.05}{\frac{1}{2}(0.05) + \frac{1}{2}(0.0025)} = 0.95$

→ In a bolt factory machines A, B, C manufacture 20%, 30% & 50% of the total of their output and 6%, 3% & 2% are defective. A bolt is drawn at random and found to be defective. Find the prob's that it is manufactured from i) A ii) B iii) C

Sol let E_1, E_2, E_3 be the events that the bolts are manufactured by the machines A, B, C

$$P(E_1) = \frac{20}{100} = \frac{1}{5}$$

$$P(E_2) = \frac{30}{100} = \frac{3}{10}$$

$$P(E_3) = \frac{50}{100} = \frac{1}{2}$$

let D denote event of bolt defective.

then $P(D|E_1) = 0.06$

$P(D|E_2) = 0.03$

$P(D|E_3) = 0.02$

i) If bolt is defective, then the prob. that it is from A

$$P(E_1|D) = \frac{P(E_1)\cdot P(D|E_1)}{P(E_1)\cdot P(D|E_1) + P(E_2)\cdot P(D|E_2) + P(E_3)\cdot P(D|E_3)}$$

$$= \frac{\frac{1}{5}(0.06)}{\frac{1}{5}(0.06) + \frac{3}{10}(0.03) + \frac{1}{2}(0.02)} =$$

→ Two bolts are drawn from a box containing 4 G and 6 B bolts. Find the prob. that the second bolt is good if the first one is found to be bad.

Sol prob. that the first bolt is bad = $\frac{6}{10}$

Prob. that the bolt drawn is good = $\frac{4}{9}$

P(first is bad and second is good) = $\frac{6}{10} \cdot \frac{4}{9} = \frac{4}{15}$

→ In a factory, Machine A produces 40% of the output and machine B produces 60%. On the average, 9 items in 1000 produced by A are defective and 1 item in 250 produced by B are defective. An item drawn at random from a day's output is defective. what is the prob that it was produced by A or B.

Sol $P(A) = 40\% = 0.4$

$P(B) = 60\% = 0.6$

let E be event of producing defective item.

$P(E|A) = \frac{9}{1000}$; $P(E|B) = 1/250$

$$P(A|E) = \frac{P(A) \cdot P(E|A)}{P(A) \cdot P(E|A) + P(B) \cdot P(E|B)}$$

$$= \frac{(0.4)\left(\frac{9}{1000}\right) \cancel{+ (0.6)\left(\frac{1}{250}\right)}}{(0.4)\left(\frac{9}{1000}\right) + (0.6)\left(\frac{1}{250}\right)} = \underline{0.6}$$

$P(B|E) = \dfrac{P(B)\cdot P(E|B)}{P(A)\cdot P(E|A) + P(B)\cdot P(E|B)} = \dfrac{0.6\times 0.004}{0.006}$

$= 0.4$

Note: prob. that is produced by A or B = 1

→ A manufacture firm produces steel pipes in three plants, with daily production of 500, 1000 & 2000 units respectively. Acc to past experience, it is known that the fraction of defective outputs produced by these plants are 0.005; 0.008; 0.010. If pipe is selected and found to be defective find

(i) from which plant the pipe came

(ii) what is the prob. that it came from first plant

Sol let E_1, E_2, E_3 be the events that pipe is manufactured in plant 1, 2 or 3.

$P(E_1) = \dfrac{500}{3500}$; $P(E_2) = \dfrac{1000}{3500}$; $P(E_3) = \dfrac{2000}{3500}$

let D be event of defective pipe is drawn.

$P(D|E_1) = 0.005$; $P(D|E_2) = 0.008$; $P(D|E_3) = 0.10$

$P(E_1|B) = 5/61$; $P(E_2|B) = 16/61$; $P(E_3/B) = \dfrac{40}{61}$

Problems. (Unit I)

Introduction to Probability & Statistics.

1. What is the probability for a leap year to have 52 Sundays and 53 Mondays?

Sol: Leap year = 366 days = 52 weeks & 2 days.

The 2 days can be in one of the 7 ways:

(1) mon, Tue (3) wed, Thu (5) Fri, Sat (7) Sun, mon.

(2) Tue, wed (4) Thu, Fri (6) sat Sun

Let E = be the event of having 52 sundays & 53 mondays. (1) is the only choice because if we chose (7) we will get 53 Sundays & 53 mondays.

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

2. In a class there are 10 Boys & 5 girls. A committee of 4 students is to be selected from the class. Find the probability for the committee to contain atleast 3 girls.

Sol. atleast 3G ⇒ ≥ 3 = 3G 1B + 4G.

$$\therefore P(E) = P(3G\,1B) + P(4G) = \frac{5_{C_3} \cdot 10_{C_1}}{15_{C_4}} + \frac{4_{C_4}}{15_{C_4}} = \frac{\cancel{5}\times\cancel{4}\times 10\times\cancel{4}\times\cancel{3}\times\cancel{2}}{\cancel{2}\times\cancel{15}\times\cancel{14}\times 13\times\cancel{12}} + \frac{\cancel{4}\times\cancel{3}\times\cancel{2}}{15\times\cancel{14}\times 13\times\cancel{12}}$$

$$= \frac{20}{273} + \frac{1}{105\times 13}$$

=

3. A class consists of 6 girls & 10 boys. If a committee of 3 is chosen at random from the class, find the probability that (i) 3 boys are selected (ii) exactly 2 girls are selected (iii) atleast 2 girls are selected (iv) atmost 2 girls are selected.

Sol: Girls = 6, Boys = 10 committee = 3 Total = 16

(i) $P(3B) = \frac{10_{C_3}}{16_{C_3}}$

(ii) $P(2G) = P(2G\,1B) = \frac{6_{C_2}\times 10_{C_1}}{16_{C_3}}$

(iii) $P(\geq 2G) = P(2G\,1B) + P(3G) = \frac{6_{C_2}\times 10_{C_1}}{16_{C_3}} + \frac{6_{C_3}}{16_{C_3}}$

(iv) $P(\leq 2G) = P(0G\,1B) + P(1G\,2B) + P(2G\,1B) = \frac{10_{C_3}}{16_{C_3}} + \frac{6_{C_1}\times 10_{C_2}}{16_{C_3}} + \frac{6_{C_2}\times 10_{C_1}}{16_{C_3}}$

4. What is the probability that a card drawn at random from the pack of playing cards may be either a queen or a king.

Sol: Total = 52 cards. No. of Q cards = 4

" " K " = 4.

$\therefore P(Q \cup K) = P(Q) + P(K) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52} = \frac{2}{13}$.

5. Out of 15 items 4 are not in good condition, 4 are selected at random. Find the probability that (i) All are not good (ii) Two are not good. (iii) atleast 2 are not good (iv) at most 2 are not good.

Sol: Total = 15 (G) Good items = 11

(NG) Not Good " = 4.

To select $\boxed{4}$.

(i) $P(4NG) = \frac{^4C_4}{^{15}C_4} = \frac{4\times3\times2\times1}{15\times14\times13\times12}$

(ii) $P(2NG) = \frac{^4C_2 \times {^{11}C_2}}{^{15}C_4}$

(iii) $P(\geq 2NG) = P(2NG) + P(3NG) + P(4NG)$

$= \frac{^4C_2 \times {^{11}C_2}}{^{15}C_4} + \frac{^4C_3 \times {^{11}C_1}}{^{15}C_4} + \frac{^4C_4}{^{15}C_4}$

(iv) $P(\leq 2NG) = P(0\,NG) + P(1\,NG) + P(2NG)$

$= \frac{^{11}C_4}{^{15}C_4} + \frac{^4C_1 \times {^{11}C_3}}{^{15}C_4} + \frac{^4C_2 \times {^{11}C_2}}{^{15}C_4}$

6. A card is drawn from a well shuffled pack of cards. What is the probability that it is either a spade or an ace?

Sol: Total cards = 52, No. of spade cards = 13, No. of ace cards = 4

We have No. of space ace card = 1

$\therefore P(S \text{ or } A) = P(S) + P(A) - P(S\cap A) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52}$

7. If $P(A\cap B) = \frac{1}{6}$, $P(A) = \frac{1}{2}$, then $P(\frac{B}{A}) = ?$

$$P(\frac{B}{A}) = \frac{P(A\cap B)}{P(A)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

8. A class has 10 boys & 5 girls. 3 students are selected at random one after another. Find the probability that (i) First 2 are boys & third is girl (ii) First & 3rd are of same gender & 2nd is of opposite gender.

Sol: 10 Boys 5 Girls Total = 15 ;

To select $\boxed{3}$.

(i) $P(BBG) = \frac{10}{15} \times \frac{9}{14} \times \frac{5}{13}$

(ii) $P(BGB) + P(GBG)$

$$= \frac{10}{15} \times \frac{5}{14} \times \frac{9}{13} + \frac{5}{15} \times \frac{10}{14} \times \frac{4}{13}.$$

9. Three machines I, II, III produce 40%, 30%, 30% of the total number of items of factory. The percentage of defective items of these machines are 4%, 2%, 3%. If an item is selected at random, find the probability that the item is defective.

Sol: Given $P(A) = 40\% = 0.4$; $P(B) = 30\% = 0.3$;

$P(C) = 30\% = 0.3$.

Let D - Defective items.

$P(D/A) = 4\% = 0.04$; $P(D/B) = 2\% = 0.02$;

$P(D/C) = 3\% = 0.03$.

$$P(D) = P(A)P(\frac{D}{A}) + P(B)P(\frac{D}{B}) + P(C)P(\frac{D}{C})$$

$$= (0.4)(0.04) + (0.3)(0.02) + (0.3)(0.03)$$

$$= \frac{31}{1000}$$

10. In a bolt factory machines A, B, C manufacture 20%, 30% and 50% of the total of their output and 6%, 3% & 2% are defective. A bolt is drawn at random & found to be defective. Find the probabilities that it is manufactured from (i) Machine A (ii) Machine B (iii) Machine C.

Sol: $P(A) = 20\% = 0.2$ D - Defective

$P(B) = 30\% = 0.3$ $P(D/A) = \frac{6}{100}$; $P(\frac{D}{B}) = \frac{3}{100}$; $P(\frac{D}{C}) = \frac{2}{100}$

$P(C) = 50\% = 0.5$

(i) $P(A/D) = \frac{P(A \cap D)}{P(D)} = \frac{P(A) \cdot P(D/A)}{P(A)P(D/A) + P(B)P(D/B) + P(C)P(D/C)}$

$= \frac{12}{31}$

(ii) $P(B/D) = \frac{P(B \cap D)}{P(D)} = \frac{P(B) \cdot P(D/B)}{P(A)P(D/A) + P(B)P(D/B) + P(C)P(D/C)}$

$= \frac{9}{31}$

(iii) $P(C/D) = \frac{P(C \cap D)}{P(D)} = \frac{P(C) \cdot P(D/C)}{P(A)P(D/A) + P(B)P(D/B) + P(C)P(D/C)}$.

$= \frac{10}{31}$

Of the 3 men, the chances that a politician, businessman or an academician will be appointed as a Vice Chancellor (V.C) of a University are 0.5, 0.3, 0.2 respectively. Probability that research is promoted by these persons 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.

(ii) If research is promoted, what is the probability that VC is an academician?

$P(A) = 0.5$, $P(B) = 0.3$, $P(C) = 0.2$, $P(R/A) = 0.3$,

$P(R/B) = 0.7$, $P(R/C) = 0.8$

(i) $P(R) = P(A)P(R/A) + P(B)P(R/B) + P(C)P(R/C)$

$= 0.52$

(ii) $P(C/R) = \frac{P(C \cap R)}{P(R)} = \frac{4}{13} = 0.307 \left\{ = \frac{P(C) \cdot P(R/C)}{P(A)P(R/A) + P(B)P(R/B) + P(C)P(R/C)} \right\}$

12. 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 plumbings. What is the probability that business man's room having faulty plumbing is assigned to hotel Z ?

Sol: $P(X) = 20\% = 0.2$; $P(F/Y) = 4\% = 0.04$

$P(Y) = 50\% = 0.5$ $\quad P(F/Z) = 8\% = 0.08$

$P(Z) = 30\% = 0.3$

$P(F/X) = 5\% = 0.05$

$$\therefore P(Z/F) = ? = \frac{P(F \cap Z)}{P(F)} = \frac{P(Z)\,P(F/Z)}{P(X)\,P(F/X) + P(Y)P(F/Y) + P(Z)P(F/Z)}$$

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

3. Suppose 5 men out of 100 and 25 women out of 10,000 are colour blind. A colour blind person is chosen at random. What is the probability of the person being a man (male) (Assume men (male) & women (female) to be in equal number.)

: No. of men being colour blind $\Rightarrow P(B/M) = \frac{5}{100} = 0.05$

" " women being " " $\Rightarrow P(B/W) = \frac{25}{10,000} = 0.00025$

(B – Colour blind say)

Given $P(M) = \frac{1}{2} = P(W)$ {$\because$ they are in equal no.s}.

$$P(M/B) = ? = \frac{P(M \cap B)}{P(B)} = \frac{P(M) \cdot P(B/M)}{P(M)\,P(B/M) + P(W)\,P(B/W)}$$

$$= \frac{(0.5)(0.05)}{(0.5)(0.05) + (0.5)(0.00025)} = 0.95$$

The chance that doctor A will diagnose a disease 'x' correctly is 60%. The chance that a patient will die by his treatment after correct diagnosis is 40% and the chance of death by wrong diagnosis is 70%. A patient of doctor A, who had disease x, died. What is the chance that his disease was diagnosed correctly.

x – disease ; C – correct diagnosis $\quad$ W – Wrong diagnosis

$P(C) = 60\% = 0.6 \therefore P(W) = 1 - P(C) = 40\% = 0.4$

$P(x/C) = 40\% = 0.4$, $P(x/W) = 70\% = 0.7$ $\quad P(C/x) = ?$

$$P(C/x) = \frac{P(C \cap x)}{P(x)} = \frac{P(C) \cdot P(x/C)}{P(C)\,P(x/C) + P(W)P(x/W)} = \frac{6}{13}$$

15. A bag A contains 2 white and 3 red balls and a bag B contains 4 white and 5 red balls. One ball is drawn at random from one of the bags and it is found to be red. Find the probability that the red ball drawn is from bag B.

Sol:

bag A — 2 W ; 3 R Total balls in bag A = 5 } Tot
bag B — 4 W ; 5 R " " " " B = 9 }

We have $P(A) = P(B) = \frac{1}{2}$ {∵ Both have equal chances of being chosen}.

Also $P(R/A) = \frac{3}{5}$; $P(R/B) = \frac{5}{9}$

To find $P(B/R) = ?$

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

$$= \frac{\frac{1}{2} \times \frac{5}{9}}{\frac{1}{2} \times \frac{3}{5} + \frac{1}{2} \times \frac{5}{9}}$$

$$= \frac{25}{52}.$$

Random Variable

A real variable 'X' whose value is determined by the outcome of a random experiment. It is also called a stochastic variable.

eg: 1. Tossing of a coin twice

Sample Space $S = \{s_1, s_2, s_3, s_4\}$

where $s_1 = HH$, $s_2 = HT$, $s_3 = TH$, $s_4 = TT$

Define a function $X : S \to R$ by $X(s) =$ 'no. of heads' then

$X(s_1) = 2$, $X(s_2) = 1$, $X(s_3) = 1$, $X(s_4) = 0$.

$\therefore$ Range $= \{2, 1, 0\}$ $\Rightarrow$ Random variable $X(s) = \{0, 1, 2\}$.

eg: 2: Throw of pair of dice

sample space $S = 6^2 = 36 = \begin{Bmatrix} (1,1), & (1,2), & \cdots & (1,6) \\ (2,1), & (2,2) & \cdots & (2,6) \\ \cdots & \cdots & \cdots & \cdots \\ (6,1), & (6,2), & \cdots & (6,6) \end{Bmatrix}$

Define a function $X : S \to R$ by

$X(S) =$ Sum on two dice, $X(i, j) = i + j \; \forall (i, j) \in S$

then $X(S) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

Types of Random variables

They are of 2 types: Discrete Random variables; Continuous Random variables.

Discrete Random variable

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

(or)

If the r.v take the values only in the set $\{0, 1, 2, \ldots, n\}$ is called discrete random variable.

eg: 1. Throw of a dice

2. Tossing of a coin

3. No. of defectives in a sample of electric bulbs etc.

Continuous Random Variable:

A random variable X which can take values continuously i.e which takes all possible values in a given interval.

eg 1. The height, age and weight of individuals

2. Temperature and time etc

Probability function of a Discrete random variable :

If for a discrete r.v 'X', the real valued function on $P(x)$ is such that $P(X=x) = p(x)$ then $p(x)$ is called Probability function or probability mass function of a disrete r.v 'X'.

Probability Distribution function.

P.D.F associated with X is the probability that the outcome of an experiment will be one of the outcomes for which $X(s) \le x, x \in R$

ie $F_X(x) = P(X \le x) = P\{s : X(s) \le x\}, \; -\infty < x < \infty$

is called Distribution function.

Properties of Distribution function

1. If F distribution function of X, and if $a < b \Rightarrow P(a < X \le b) = F(b) - F(a)$
2. $P(a \le X \le b) = P(X=a) + [F(b) - F(a)]$
3. $P(a < X < b) = [F(b) - F(a)] - P(X=b)$
4. $P(a \le X < b) = [F(b) - F(a)] - P(X=b) + P(X=a)$

NOTE: $0 \le F(x) \le 1$.

Discrete Probability Distribution (Probability mass function)

It is the set of its possible values together with their respective probabilities.

Let X be a discrete r.v with possible outcomes $x_1, x_2, x_3, \ldots$ then their probabilities $p_i = P(X = x_i) = p(x_i)$ for $i = 1, 2, 3, \ldots$

if $p(x_i) > 0$ & $\sum_{i=1}^{n} p(x_i) = 1$ then the function "$p(x)$" is called the 'probability mass function' of r.v X & $\{x_i, p(x_i)\}$ for $i = 1, 2, \ldots$ is called Discrete Probability Distribution.

Eg: Tossing a Coin two times with random variable $X(s)$ = no. of heads $= \{0,1,2\}$

then $P(X=0) = \frac{1}{4}$; $P(X=1) = \frac{1}{2}$, $P(X=2) = \frac{1}{4}$

x_i	0	1	2
$p(x_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$\Rightarrow \sum p(x_i) = 1$ & $p(x_i) > 0$.

Thus total probability '1' is distributed into 3 parts as $\frac{1}{4}, \frac{1}{2}$ & $\frac{1}{4}$.

Note: 1) Frequency distribution tells how total frequency distributed among different values of the variable.

2) Probability distribution tells how total probability '1' is distributed among the values which the random variable can take.

Cumulative Distribution Function of a Discrete R.V.

$$F(x) = P(X \le x) = \sum_{i=1}^{x} P(x_i) \text{ where } x \text{ is any integer}$$

then $F(x)$ is called cumulative Distribution function of X.

i.e $F(x) = \begin{cases} 0, & -\infty < x < x_1 \\ P(x_1), & x_1 \le x < x_2 \\ P(x_2), & x_2 \le x < x_3 \\ \ldots \\ \ldots \\ P(x_1) + P(x_2) + \cdots + P(x_n), & x_n \le x < \infty \end{cases}$

Expectation of Discrete Probability Distribution

Let X be a r.v with $x_1, x_2, \ldots x_n$ with probabilities $p_1, p_2, \ldots, p_n$ then

Mathematical Expectation or Expected value of X is defined as "Sum of products of different values of x & their corresponding probabilities".

$$\boxed{\text{i.e } E(X) = \sum_{i=1}^{n} x_i p_i},$$

In General $\boxed{E(g(x)) = \sum_{i=1}^{n} p_i\, g(x_i)}$

Results:
1. $E(X) = \mu$ $\left\{\because \text{Mean} = \frac{\sum p_i x_i}{\sum p_i} = \sum p_i x_i \text{ as } \sum p_i = 1\right\}$
2. $E(X+k) = E(X) + k$ (k is a constant)
3. $E(kX) = k\,E(X)$
4. $E(aX \pm b) = aE(X) \pm b$
5. $E(X+Y) = E(X) + E(Y)$ provided $E(X), E(Y)$ exists.
6. $E(X - \bar{X}) = 0$
7. $E(XY) = E(X) \cdot E(Y)$ if X, Y are two independent r.vs.
8. $E\left(\frac{1}{X}\right) \ne \frac{1}{E(X)}$

Mean: The mean 'μ' of the Distribution function is given by

$$\boxed{\mu = \frac{\sum p_i x_i}{\sum p_i} = \sum p_i x_i}$$

Variance : Variance of the probability distribution of r.v X is defined as

$$\boxed{Var(X) = E[X - E(X)]^2}$$

Since variance of $X = V(X) = \sum_{i=1}^{n} (x_i - \mu)^2 p_i = \sum_{i=1}^{n} p_i x_i^2 - \mu^2$

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

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

Standard Deviation $S.D = \sigma = \sqrt{Var(X)} = \sqrt{E[(X)-(E(X)]^2}$

Results: 1. $Var(K) = 0$ where k is constant $\left\{\begin{array}{l} \because \text{mean} = \text{constant} \\ \Rightarrow X - \mu = 0 \end{array}\right\}$

2. $V(KX) = K^2 V(X)$

3. $V(X+K) = V(X)$

★ 4 $V(aX+b) = a^2 V(X)$ where a, b are constants.

Sol: Let $Y = aX+b$

$\therefore E(Y) = E(aX+b)$

$= aE(X) + b$

Consider $Y - E(Y) = aX + \cancel{b} - aE(X) - \cancel{b} = a[X - E(X)]$

Squaring & taking expectation, we get

$E[Y - E(Y)]^2 = a^2 E(X - E(X))^2$

$\Rightarrow Var(Y) = a^2 Var(X)$

$\Rightarrow V(aX+b) = a^2 V(X)$.

Hence Proved.

Do Problems.

Continuous Probability Distribution: (Probability Density function)

Def^n: By considering small interval $[x - \frac{dx}{2}, x + \frac{dx}{2}]$ of length dx round the point x. Let $f(x)$ be any continuous function of x so that $f(x)dx$ represents the probability that the variable X falls in that interval. ie $P(x - \frac{dx}{2} \le X \le x + \frac{dx}{2}) = f(x)dx$ then $f(x)$ is called the probability density function & the curve $y = f(x)$ is known as probability density curve.

The probability for a variate value to fall in the finite interval (a, b) is $\int_a^b f(x)dx$ which represents the area b/w the curve $y = f(x)$, x-axis & the ordinated $x = a$ & $x = b$.

properties of density function

(i) $f(x) \geq 0 \ \forall x \in R$

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

(iii) $P(E) = \int_E f(x)dx$ is well defined for any event E.

★ Note: Since continuous variable associate with intervals, the probability at a particular point is always zero. Thus $P(a<X<b) = P(a \le X \le b)$
$P(a < X \le b) = P(a \le X < b) = F(b) - F(a)$

Cumulative Distribution function of a Continuous Random Variable :

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

Probability density function (p.d.f) $\boxed{f_X(x) = \frac{d}{dx} F_X(x)}$

Theoretical Distributions (or Probability Distributions) are mathematical models. In this, variates are distributed as per some definite law which can be expressed mathematically. A theoretical distribution is the frequency distribution [w.r.t. frequency distributions are based on actual observations or experiment] of a certain event in which frequencies are obtained by mathematical computation. They give a very close approximation to statistical measures such as mean, variance, standard deviation etc, as calculated from actual frequency distribution.

There are two types of theoretical distributions.

(i) Discrete Theoretical Distributions

ⓐ Binomial Distribution
ⓑ Poisson "
ⓒ Rectangular "
ⓓ Negative Binomial "
ⓔ Geometric ".

(ii) Continuous Theoretical Distributions

ⓐ Normal distribution
ⓑ Student's t- "
ⓒ Chi-Square "
ⓓ F- distribution.

Binomial Distribution.

It was discovered by James Bernoulli in the year 1700 & is a discrete probability distribution.

The conditions for the applicability of a B.D are :

(i) There are 'n' independent trials (ie trials are repeated under identical conditions)

(ii) There are only 2 possible outcoms for each trial. Sucess 'p' & failure 'q'.

(iii) The trials are independent, ie the probability of an event in any trial is not affected by the results of any other trial.

(iv) The probability of success in each trial remains constant & does not change from trial to trial.

Def^n: A r.v. X has a B.D if it assumes only non-negative values & its probability density function is given by

$$P(X=r) = p(r) = \begin{cases} {}^nC_r p^r q^{n-r} & ; r=0,1,2,\dots,n \; ; q=1-p \\ 0 & , \text{otherwise} \end{cases}$$

$$= b(r; n, p)$$

Here n, p are called parameters of the distribution as they are the 2 independent constants. 'n' is sometime known as "degree of the distribution".

egs of B.D : (i) No. of defective bolts in a box containing 'n' bolts.

(ii) No. of post graduates in a group of 'n' men, etc.

The Binomial Distribution function is:

$$F_X(x) = P(X \le x) = \sum_{r=0}^{n} {}^nC_r p^r q^{n-r}.$$

Binomial Frequency Distribution: The possible no. of successe and their frequencies is called a Binomial frequency Distribution.

If 'n' independent trials constitute one experiment and this expt is repeated 'N' times, then the frequency of 'r' successes is $\underline{N \cdot {}^nC_r\, p^r q^{n-r}}$.

$\therefore$ In 'N' sets of 'n' trials the theoretical frequencies of $0, 1, 2, \ldots, r, \ldots, n$ successes are given by the terms of expansion of $\underline{N(q+p)^n}$.

Note ① The probabilities of $0, 1, 2, \ldots, r, \ldots, n$ successes in 'n' trials are given by the terms of the binomial expansion of $(q+p)^n$.

ie $(q+p)^n = q^n + {}^nC_1\, q^{n-1} p + {}^nC_2\, q^{n-2} p^2 + \cdots + {}^nC_r\, q^{n-r} p^r + \cdots + p^n$.

Here the probability of exactly 'r' successes is ${}^nC_r\, p^r q^{n-r}$.

② The probability of $\underline{\text{no successes}}$ in n trials is $= q^n$

" " " $\underline{\text{all successes}}$ " " " $= p^n$

" " " $\underline{\text{atleast One success}}$ " " " $= {}^nC_1\, q^{n-1} p + {}^nC_2\, q^{n-2} p^2 + \cdots + p^n = \underline{\underline{1-q^n}}$.

~~Do problems~~.

Constants of B.D:

(1) Mean of X : $\mu = E(X) = \sum_{r=0}^{n} r\, p(r) = \sum_{r=0}^{n} r \cdot {}^nC_r\, p^r q^{n-r}$

$$\Rightarrow \mu = 1 \cdot {}^nC_1\, p q^{n-1} + 2\, {}^nC_2\, p^2 q^{n-2} + \cdots + n \cdot {}^nC_n\, q^{n-n} p^n$$

$$= n p q^{n-1} + 2\frac{n(n-1)}{2!} p^2 q^{n-2} + \cdots + n p^n$$

$$= np\left[q^{n-1} + \cancel{2}\frac{(n-1)}{\cancel{2!}} p q^{n-2} + \cancel{3}\frac{(n-1)(n-2)}{\cancel{3}(2)} p^2 q^{n-3} + \cdots + p^{n-1}\right]$$

$$= np\left[q^{n-1} + (n-1) p q^{n-2} + \frac{(n-1)(n-2)}{2!} p^2 q^{n-3} + \cdots + p^{n-1}\right]$$

$$= np\left[{}^{n-1}C_0\, q^{n-1} p^0 + {}^{n-1}C_1\, p^1 q^{n-2} + {}^{n-1}C_2\, p^2 q^{n-3} + \cdots + {}^{n-1}C_{n-1}\, p^{n-1}\right]$$

$$= np\,[q+p]^{n-1}$$

$$\boxed{\therefore \mu = np}$$

2. Variance of B.D:

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

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

$$= \sum_{r=0}^{n} r^2 p(r) - \left[\sum_{r=0}^{n} r\, p(r)\right]^2$$

$$= \sum_{r=0}^{n} [r(r-1)+r]p(r) - \mu^2$$

$$= \sum_{r=0}^{n} r(r-1)p(r) + \sum_{r=0}^{n} r\, p(r) - \mu^2$$

$$= \sum_{r=0}^{n} r(r-1)p(r) + \mu - \mu^2$$

$$= 2(2-1)\,{}^{n}C_2\, p^2 q^{n-2} + 3(3-1)\,{}^{n}C_3\, p^3 q^{n-3} + \cdots + n(n-1)\,{}^{n}C_n\, p^n q^{n-n} + \mu - \mu^2$$

$$= 2\,{}^{n}C_2\, p^2 q^{n-2} + 6\,{}^{n}C_3\, p^3 q^{n-3} + \cdots + n(n-1)p^n + \mu - \mu^2$$

$$= \cancel{2}\,\frac{n(n-1)}{\cancel{2}} p^2 q^{n-2} + \cancel{6}\,\frac{n(n-1)(n-2)}{\cancel{3\times 2}} p^3 q^{n-3} + \cdots + n(n-1)p^n + \mu - \mu^2$$

$$= n(n-1)p^2\left[q^{n-2} + (n-2)pq^{n-3} + \cdots + p^{n-2}\right] + \mu - \mu^2$$

$$= n(n-1)p^2\left[{}^{n-2}C_0\, q^{n-2} + {}^{n-2}C_1\, p^1 q^{n-3} + \cdots + {}^{n-2}C_{n-2}\, p^{n-2}\right] + \mu - \mu^2$$

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

$$= n(n-1)p^2 \quad (\because q+p=1) + \mu - \mu^2$$

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

$$= np(1-p)$$

$$= npq \quad \{\because p+q=1 \Rightarrow q = 1-p\}$$

$$\boxed{\therefore Var(X) = \sigma^2 = npq}$$

Poisson Distribution

S.D. Poisson (Simeon Denis Poisson) (1837) introduced Poisson Distribution as a rare distribution of rare events. i.e the events whose probability of occurrence is very small but the no. of trials which could lead to the occurrence of the Event, are very large.

Conditions of P.D : (i) The no. of trials 'n' is large.
(ii) The probability of success 'p' is very small.
(iii) $np = \lambda$ is finite.

Examples of P.D :
(1) The no. of printing mistakes per page in a large text.
(2) The no. of cars passing a certain point in 1 minute.

<u>Definition</u>: A r.v X is said to follow P.D if it assumes only non-negative values & its probability density function is given by

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

Here $\lambda > 0$ is called <u>parameter</u> of the distribution.

Constants of Poisson Distribution :

(1) Mean $= \mu = E(X) = \lambda$

Proof : $E(X) = \sum x\, p(x) = \sum_{x=0}^{\infty} \dfrac{x e^{-\lambda}\lambda^{x}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \dfrac{x \lambda^{x}}{x!}$

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

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

~~Take x-1=y~~ $= \lambda e^{-\lambda}\left[1 + \lambda + \dfrac{\lambda^2}{2!} + \dfrac{\lambda^3}{3!} + \cdots\right]$

$$= \lambda e^{-\lambda}\left[e^{\lambda}\right]$$

$$= \lambda$$

$$\boxed{\therefore \mu = \lambda}$$

(2) Variance of P.D

$V(X) = \lambda$

$= E(x^2) - [E(x)]^2 = E(x^2) - \mu^2 = E(x^2) - \lambda^2$.

Proof: $Var(X) = E(x^2) - [E(x)]^2$

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

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

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

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

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

$$= \lambda^2 e^{-\lambda}(e^{\lambda}) + \lambda e^{-\lambda}(e^{\lambda})$$

$$= \lambda^2 e^0 + \lambda e^0 = \lambda^2 + \lambda \Rightarrow E(x^2) = \lambda^2 + \lambda$$

$$\therefore Var(X) = E(x^2) - \lambda^2 = \cancel{\lambda^2} + \lambda - \cancel{\lambda^2} = \lambda \Rightarrow \boxed{Var(X) = \lambda}$$

Note. S.D $\sigma = \sqrt{\lambda}$.

Recurrence Relation:

$$p(x+1) = \frac{\lambda}{x+1} p(x).$$

Note: When n is large say greater than 30 & p is very small say less than 0.1 then B.D can be approximated by P.D.

Normal Distribution.

It is a continuous distribution. N.D was first discovered by English Mathematician De-Moivre (1667-1745) in 1733 & further refined by French Mathematician Laplace (1749-1827) in 1774 & independently by Karl Friedrich Gauss (1777-185).

It is also known as Gaussian Distribution.

It is another limiting form of Binomial Distribution. It is found so often in real life that it is called Normal Distribution, the name which is commonly used.

Defn A r.v X is said to have a N.D if its p.d.f is given by $f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, $-\infty<x<\infty$, $\sigma>0$, $-\infty<\mu<\infty$.

μ is Mean & σ is Standard deviation of x.

μ, σ are called parameters of N.D.

A linear combination of independent normal variates is also a normal variate.

Chief Characteristics of N.D:

1. The graph of Normal Distribution in xy-plane is known as normal curve.
2. The curve is a bell-shaped curve & symmetrical w.r.t mean 'μ'. ie The two tails on right & left sides of the mean 'μ' extends to infinity.
3. Area under the normal curve represents total population.
4. Normal curve is Unimodal (ie has only one max. pt.) At $x=\mu$: mean = Median = Mode (as distribution is symmetrical).
5. x-axis is an asymptote to the curve.
6. The points of inflexion of the curve are at $x=\mu\pm\sigma$ & curve changes from concave to convex at $x=\mu+\sigma$ to $x=\mu-\sigma$.

7. Area under the normal curve is distributed as:

(i) Area of normal curve b/w $\mu-\sigma$ & $\mu+\sigma$ is 68.27%.

(ii) " " " " " $\mu-2\sigma$ & $\mu+2\sigma$ is 95.43%

(iii) " " " " " $\mu-3\sigma$ & $\mu+3\sigma$ is 99.73%.

Note: The total area bounded by the curve & x-axis is <u>One</u>.

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

$P(a<x<b)$ = Area under normal curve b/w the vertical lines $x=a$ & $x=b$.

$= \int_a^b f(x)\,dx$.

Multiple Random Variables.

In many practical problems several r.vs interact with each other. Multiple R.Vs help us to determine the joint statistical properties like mean, variance etc.

We study about 2-dimensional r.vs, which can be easily extended to multiple r.vs.

A 2-dimensional r.v is denoted by (X,Y) where the random vector (X,Y) is outcome of a trial which occurs in pairs. ie $X=x$, $Y=y$.

Def^n: An 'n' dimensional random vector (vector of r.vs) is a function from sample space S into R^n (n dimensional Euclidean Space)

Multiple r.vs are also of 2 types: (1) Discrete & (2) Continuous.

It is called Discrete if (X,Y) assume only finite no. of pairs.

Discrete Multiple Random Variables:

Joint Probability Mass function (Joint Probability function)

If (X,Y) is a discrete 2-dimensional r.v then the function $f(x,y)$ from R^2 into R is called JPMF & is given by
$$\boxed{f(x,y) = P(X=x, Y=y) = P(x_i, y_j)}$$

Note: (a) $0 \leq P(x_i, y_j) \leq 1$. (b) $\sum_i \sum_j P(x_i, y_j) = 1$.

Properties: (1) $P(X=x_i) = p(x_i) = \sum_j P(x_i, y_j)$

(2) $P(Y=y_j) = p(y_j) = \sum_i P(x_i, y_j)$

(3) $P(x_i) \geq P(x_i, y_j)$ for any j

(4) $P(y_j) \geq P(x_i, y_j)$ for any i

Joint Cumulative Distribution Function.

Def^n: The joint probability on cumulative distribution function of 2 r.v.s uniquely defines the probability of the joint events $\{X \le x, Y \le y\}$. It is defined by

$$\boxed{F_{XY}(x,y) = P(X \le x, Y \le y)} = \sum_{i=0}^{x_i} \sum_{j=0}^{y_j} P(x_i, y_j)$$

Properties:

① $0 \le F_{XY}(x,y) \le 1$

② $F_{XY}(-\infty, \infty) = 1$.

③ $F_{XY}(-\infty, y) = F_{XY}(x, \infty) = 0$.

* ④ F_{XY} is non-decreasing.

* ⑤ $F_X(x) = F_{XY}(x, \infty) = P(X \le x) = P(X \le x, Y \le \infty)$.

⑥ $F_Y(x) = F_{XY}(\infty, y) = P(Y \le y) = P(X \le \infty, Y \le y)$.

Marginal Probability Distribution Function.

$\boxed{F_{XY}(x,\infty), F_{XY}(\infty, y)}$ are called marginal PDF of X & Y respectively. It is given by

$$\boxed{\begin{aligned} F_{XY}(x,\infty) &= \sum_j P(x, y_j) \\ F_{XY}(\infty, y) &= \sum_i P(x_i, y). \end{aligned}}$$

eg: Spse 2 dice are rolled simultaneous, then probability of getting 3 on first die is ?

$P(X=3) = P_{XY}(3,\infty) = P(X=3, Y=1) + P(X=3, Y=2) + P(X=3, Y=3) + P(X=3, Y=4) + P(X=3, Y=5) + P(X=3, Y=6)$

$= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36}$

$= \frac{1}{6}$

Continuous Multiple Random Variables:

Joint Probability density function: It is denoted by $\underline{f_{XY}(x,y)}$. Also $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$

$$f_{XY}(x,y)\,dx\,dy = P(x \leq X \leq (x+dx),\ y \leq Y \leq (y+dy))$$

Note: $\boxed{f_{XY}(x,y) = \dfrac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}}$ Relation b/w JPF & JPDF.

Properties:

① $P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{XY}(x,y)\,dx\,dy$

② $F_{XY}(x_1,y_1) = \int_{-\infty}^{x_1}\int_{-\infty}^{y_1} f_{XY}(x,y)\,dx\,dy$

Marginal Probability density function (MPDF)

The MPDF of one of the r.vs is the integral of joint p.d.f over the other r.v.

$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$ is marginal p.d.f of X.

$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$ is " " " Y.

Statistical properties of Jointly Distributed Random Variables.

Def^n: 2 jointly distributed r.vs X, Y are said to be statistically independent of each other iff

$$f_{XY}(x,y) = f_X(x) f_Y(y) \quad \forall (x,y) \text{ in the given range}$$

ie joint p.d.f = product of 2 marginal p.d.fs.

Also if $F_{X_1X_2}(x_1,x_2) = F_X(x_1) \cdot F_X(x_2) \ \forall\ x_1, x_2$ then also the 2 r.vs X_1, X_2 are said to be statistically independent.

Conditional Probability ~~distribution~~ Density function.

Def^n:
$$f_Y(Y/X) = \frac{f_{XY}(x,y)}{f_X(x)} \quad \text{is conditional p.d.f of } Y.$$

$$f_X(X/Y) = \frac{f_{XY}(x,y)}{f_Y(y)} \quad \text{is conditional p.d.f of } X.$$

Note: If X, Y are statistically independent Then

$$f_X(X/Y) = f_X(x)$$
$$\& \quad f_Y(Y/X) = f_Y(y)$$

$$\because f_X(X/Y) = \frac{f_{XY}(x,y)}{f_Y(y)} \quad \text{(by def}^n\text{)}$$
$$= \frac{f_X(x) \cdot \cancel{f_Y(y)}}{\cancel{f_Y(y)}} \quad \{\because \text{Statistical independence}$$
$$= f_X(x)$$

Conditional Probability Distribution Function.

Def^n: Let X, Y be 2 r.vs discrete or continuous.

$$F_Y(Y/X) = \frac{F_{XY}(x,y)}{F_X(x)}, \quad F_X(x) > 0 \text{ is conditional}$$

distribution of r.v Y given $X = x$.

Also $$F_X(X/Y) = \frac{F_{XY}(x,y)}{F_Y(y)}, \quad F_Y(y) > 0 \text{ is conditional}$$

distribution of r.vs X given $Y = y$.

Problems related to Binomial Distribution.

① Pg 125 Pb 3. A die is thrown 6 times. If getting an even no. is a success, find the probabilities of (i) atleast one success (ii) ≤ 3 successes (iii) 4 successes.

w.k.t $P(x=r) = {}^nc_r p^r q^{n-r}$.

Sol: Given $n = 6$. even no. is success

p = getting an even no. on die.

we have 2, 4, 6 as even no.s on die.

$\therefore p = \frac{3}{6} = \frac{1}{2} \Rightarrow q = \frac{1}{2}$ ($\because$ Total faces of dice = 6, Even " " " = 3)

(i) $P(X \geq 1) = 1 - P(X=0)$
$= 1 - {}^6c_0 p^0 q^{6-0} = 1 - q^6 = 1 - \left(\frac{1}{2}\right)^6 = \frac{63}{64}$

(ii) $P(X \leq 3) = P(x=0) + P(x=1) + P(x=2) + P(x=3)$
$= {}^6c_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^6 + {}^6c_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^5 + {}^6c_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^4 + {}^6c_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^3$
$= \left(\frac{1}{2}\right)^6 \left[1 + 6 + \frac{\cancel{6}^{3} \times 5}{\cancel{2}} + \frac{\cancel{6} \times 5 \times 4}{\cancel{3 \times 2}}\right]$
$= \frac{1}{64}[1 + 6 + 15 + 20] = \frac{\cancel{42}^{21}}{\cancel{64}_{32}} = \frac{21}{32}$.

(iii) $P(x=4) = {}^6c_4 p^4 q^{6-4} = \frac{\cancel{6}^{3} \times 5}{\cancel{2}} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^2 = \frac{15}{64}$

② Pg 126 Pb 4: 10 Coins are thrown simultaneously. Find the probability of getting atleast (i) 7 heads (ii) 6 heads.

Sol. Here $n = 10$, p = getting head $= \frac{1}{2} \Rightarrow q = \frac{1}{2}$.

(i) $P(X \geq 7) = P(x=7) + P(x=8) + P(x=9) + P(x=10)$
$= {}^{10}c_7 \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^3 + {}^{10}c_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^2 + {}^{10}c_9 \left(\frac{1}{2}\right)^9 \left(\frac{1}{2}\right) + {}^{10}c_{10} \left(\frac{1}{2}\right)^{10}$
$= \left(\frac{1}{2}\right)^{10} \left[\frac{10 \times \cancel{9}^{3} \times \cancel{8}^{4}}{\cancel{3} \times \cancel{2}} + \frac{\cancel{10}^{5} \times 9}{\cancel{2}} + 10 + 1\right]$.
$= \frac{(120 + 45 + 11)}{1024} = \frac{\cancel{176}^{\cancel{88}\ 11}}{\cancel{1024}_{\cancel{512}\ \cancel{128}\ 64}} = \frac{11}{64}$.

(ii) $P(X \geq 6) = P(x=6) + P(x \geq 7)$
$= {}^{10}c_6 \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^4 + \frac{11}{64} = \frac{10 \times \cancel{9}^{3} \times \cancel{8}^{2} \times 7}{\cancel{4} \times \cancel{3} \times \cancel{2}} \cdot \frac{1}{1024} + \frac{176}{1024}$
$= \frac{210 + 176}{1024} = \frac{\cancel{386}^{193}}{\cancel{1024}_{512}} = \frac{193}{512}$.

(3) If 3 of 20 tyres are defective & 4 of them are randomly chosen for inspection, what is the probability that only one of the defective tyre will be included? (Pg 127 Pb 7)

Sol: Given = 20 tyres; 3 are defective ⇒ 17 are good ⇒ $p = \frac{3}{20}$ is defective tyres probability.

To chose 4 at random ⇒ $n = 4$.

$P(x=1) = ?$ $\ {}^nc_1 p^1 q^{n-1}$

$= {}^4c_1 \left(\frac{3}{20}\right)\left(\frac{17}{20}\right)^3 = \frac{4 \times 3 \times 17^3}{(20)^4} = 0.3685$

→ Pg 127 Pb 7.

(4) Determine B.D for which mean is 4 & variance 3 (Pg 128 Pb 9)

B.D = $b(x; n, p)$. Give $\mu = 4$, $\sigma^2 = 3$

To find n; p

w.r.t $\mu = np \Rightarrow np = 4$ Also $\sigma^2 = npq = 3 \Rightarrow npq = 3$

$\therefore npq = 3 \Rightarrow 4q = 3 \Rightarrow q = \frac{3}{4} \Rightarrow p = 1-q = \frac{1}{4}$

$\therefore n\left(\frac{1}{4}\right) = 4 \Rightarrow n = 16.$

Thus $b(x; n, p) = \underline{\underline{b(x; 16, \tfrac{1}{4})}}$

(5) In 256 sets of 12 tosses of a coin, in how many cases one can expect 8 heads & 4 tails. (Pg 128 Pb 11)

Sol: $N = 256$, $n = 12$, p = getting head is success $= \frac{1}{2} \Rightarrow q = \frac{1}{2}$

$P(x=8) = ? = {}^{12}c_8 p^8 q^4 = {}^{12}c_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^4 = \frac{\cancel{12}^{5} \times 11 \times \cancel{10} \times 9}{\cancel{4 \times 3} \times \cancel{2}} \left(\frac{1}{2}\right)^{12}$

$= \frac{495}{(2)^{12}}$

∴ Expected no. of such cases in 256 sets

$= N \cdot P(x=8) = \frac{(256)(495)}{(2)^{12}} = \frac{2^8 (495)}{2^{12}} = \frac{495}{2^4}$

$= \frac{495}{16}$

$= 30.9375$

$\simeq 31$

6. Six dice are thrown 729 times How many times do you expect atleast three dice to show a 5 or 6? (Pg 129 Pb 13)

Sol: $N = 729$, $n = 6$, $p =$ to show 5 or 6 $= \frac{2}{6} = \frac{1}{3} \Rightarrow q = \frac{2}{3}$.

$P(x \geq 3) = ? = {}^nc_3 p^3 q^{n-3} + {}^nc_4 p^4 q^{n-4} + {}^nc_5 p^5 q^{n-5} + {}^nc_6 p^6 q^{n-6}$

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

$= {}^6c_3 \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^3 + {}^6c_4 \left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)^2 + {}^6c_5 \left(\frac{1}{3}\right)^5 \left(\frac{2}{3}\right) + {}^6c_6 \left(\frac{1}{3}\right)^6$.

$= \frac{\cancel{6} \times 5 \times 4}{\cancel{3 \times 2}} \times \frac{8}{3^6} + \frac{6 \times 5}{\cancel{2}} \times \frac{\cancel{4}^{2}}{3^6} + 6 \times \frac{2}{3^6} + \frac{1}{3^6}$

$= \frac{160 + 60 + 12 + 1}{3^6} = \frac{233}{3^6} = \frac{233}{729}$

$\therefore$ Expected no. of such cases in 729 times $= N \cdot P(x \geq 3)$

$= \cancel{729} \left(\frac{233}{\cancel{729}}\right) = \underline{\underline{233}}$.

★ 7 (i) Out of 800 families with 5 children each, how many would you expect to have (a) 3 boys (b) 5 girls (c) either 2 or 3 boys (d) atleast one boy? Assume equal probabilities for boys & girls.

(ii) Out of 800 families with 4 children each, how many families would be expected to have (a) 2 Boys & 2 girls (b) atleast one boy (c) no girl (d) atmost one girls? Assume equal probabilities for boys & girls.

(Pg 131 Pb 19)

Sol: (i) $N = 800$, $n = 5$, $p =$ probability of child to be boy $= \frac{1}{2} \Rightarrow q = \frac{1}{2}$.

(a) $P(X=3) = ? = {}^5c_3 p^3 q^2 = {}^5c_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = \frac{5 \times \cancel{4}}{2} \frac{1}{\cancel{4} \times 8} = \frac{5}{16}$.

(b) $P(x=0) = {}^5c_0 p^0 q^5 = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$.

(c) $P(x=2) + P(x=3) = {}^5c_2 p^2 q^3 + {}^5c_3 p^3 q^2 = \frac{5 \times \cancel{4}^{2}}{\cancel{2}} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 + \frac{5 \times \cancel{4}^{2}}{\cancel{2}} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2$

$= \frac{\cancel{20}5}{4 \times 8} = \frac{5}{8}$.

(d) $P(x \geq 1) = 1 - P(x=0) = 1 - \frac{1}{32} = \frac{31}{32}$.

Thus (a) For 800 families Expectation of 3 boys $= 800 \cdot P(x=3)$

$\Rightarrow \cancel{800}^{50} \times \frac{5}{\cancel{16}} = \underline{250}$ families

(b) Expected no. of families to have 5 girls $= N \cdot P(x=0)$

$= \cancel{800}^{\cancel{100}\,25} \times \frac{1}{\cancel{32}_{4}} = \underline{25}$ families.

(c) Expected no. of families to have either 2 or 3 boys

$= N \cdot [P(x=2) + P(x=3)] = \cancel{800}^{100} \times \frac{5}{\cancel{8}} = \underline{500}$ families.

(d) Expected no. of families with atleast one boy

$= N \cdot P(x \geq 1) = \cancel{800}^{25} \times \frac{31}{\cancel{32}} = \underline{775}$ families.

(ii) $N = 800$, $n = 4$, p = probability of child to be boy $= \frac{1}{2} \Rightarrow q = \frac{1}{2}$.

(a) $P(X=2) = {}^nC_2 p^2 q^2 = {}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = \frac{\cancel{4} \times 3}{2} \times \frac{1}{\cancel{4} \times 4} = \frac{3}{8}$

∴ Expected no. of families to have 2 Boys & 2 girls

$= N \cdot P(x=2) = \cancel{800}^{100} \times \frac{3}{\cancel{8}} = \underline{300}$ families.

(b) $P(x \geq 1) = 1 - P(x=0) = 1 - {}^4C_0 p^0 q^4 = 1 - \left(\frac{1}{2}\right)^4 = \frac{15}{16}$

∴ Expected no. of families to have atleast 1 boy

$= N \cdot P(\cancel{X} \geq 1) = \cancel{800}^{50} \times \frac{15}{\cancel{16}} = \underline{750}$ families.

(c) $P(X=4) = {}^4C_4 \left(\frac{1}{2}\right)^4 = \frac{1}{16}$ {∵ No girl ⇒ all 4 are boys}

∴ Expected no. of families to have ~~atleast~~ ~~all~~ NO girls

$= N \cdot P(X=4) = \cancel{800}^{50} \times \frac{1}{\cancel{16}} = \underline{50}$ families.

(d) $P(X=4) + P(X=3)$ (∵ Atmost 1 girl ⇒ No girl or 1 girl ⇒ 4 Boys or 3 boys)

$= {}^4C_4 \left(\frac{1}{2}\right)^4 + {}^4C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)$

$= \frac{1}{16} + 4 \times \frac{1}{16} = \frac{5}{16}$

∴ Expected no. of families to have atmost 1 girl

$= N \cdot [P(X=3) + P(X=4)]$

$= \cancel{800}^{50} \times \frac{5}{\cancel{16}} =$

$= \underline{250}$ families

⑧ (Pg 135 Pb25) In a B.D consisting of 5 independent trials, probabilities of 1 & 2 success are 0.4096 & 0.2048 respectively. Find the parameter 'p' of the distribution.

Sol: $n = 5$, $P(x=1) = 0.4096$; $P(x=2) = 0.2048$. $p = ?$

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

$\therefore P(x=1) = {}^5C_1\, p\, q^{5-1} = 0.4096 \Rightarrow 5pq^{4} = 0.4096$ — ①

$P(x=2) = {}^5C_2\, p^2 q^{5-2} = 0.2048 \Rightarrow \frac{5\times\cancel{4}^2}{\cancel{2}} p^2 q^3 = 10p^2q^3 = 0.2048$ — ②

$\frac{①}{②} \Rightarrow \frac{P(x=1)}{P(x=2)} = \frac{\cancel{5}pq^{\cancel{4}}}{\cancel{10}_2 p^{\cancel{2}} q^{\cancel{3}}} = \frac{\cancel{0.4096}^2}{\cancel{0.2048}_1}$

$\Rightarrow \frac{q}{2p} = 2 \Rightarrow q = 4p$ $\{\because p = 1-q\} \Rightarrow 1-p = 4p$
$\Rightarrow 1 = 5p$
$\Rightarrow p = \frac{1}{5}$
$\Rightarrow \boxed{p = 0.2}$

OR

By recurrence relation $P(r+1) = \frac{(n-r)p}{(r+1)q} P(r)$.

Thus $P(x=2) = \frac{(n-r)p}{(r+1)q} P(x=1)$

$= \frac{(5-1)p}{2q} P(x=1)$

$\Rightarrow \cancel{0.2048}_1 = \frac{4p}{\cancel{2}q}(\cancel{0.4096}_{\cancel{2}})$

$\Rightarrow q = 4p \Rightarrow 1-p = 4p \Rightarrow 1 = 5p \Rightarrow p = \frac{1}{5} = 0.2$

⑨ (Pg 138 Pb32) A coin is biased in a way that a head is twice as likely to occur as likely to occur as a tail. If the coin is tossed 3 times, find the probability of getting 2 tail & 1 head.

Sol: Let p = probability of getting tail as success, $n = 3$.

Given $P(H) = 2P(T)$ w.k.t $P(H) + P(T) = 1 \Rightarrow 3P(T) = 1$
$\Rightarrow P(T) = \frac{1}{3} \Rightarrow p = \frac{1}{3}$.
$\boxed{\therefore P(H) = \frac{2}{3}} \Rightarrow q = \frac{2}{3}$.

To find $P(2T) = ?$ $P(x=2) = {}^nC_2\, p^2 q^{n-2}$

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

$= \cancel{3}\left(\frac{1}{3}\right)^2 \times \frac{2}{\cancel{3}} = \frac{2}{9}$.

(10) Pg139 Pb 33. Fit a B·D to the following frequency distribution.

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

Fit a B·D ⇒ Obtain $E(x)$ where $f(x) = N \cdot p(x)$; $N = \sum_i f_i$ & $p(x) = f(x)$.

From data $n = 6$ ($\because x = 0, 1, \dots 6$) ; $N = \sum_i f_i = 13+25+52+58+32+6+4 = 200$

$$\text{Mean} = \frac{\sum_i x_i f_i}{\sum_i f_i} = np \Rightarrow \frac{(0\times13)+(1\times25)+(2\times52)+(3\times58)+(4\times32)+(5\times16)+(6\times4)}{200}$$

$$\Rightarrow np = \frac{535}{200} = 2.675$$

$\because n = 6 \Rightarrow 6p = 2.675 \Rightarrow p = 0.446 \Rightarrow q = 0.554$.

$\therefore P(x=0) = {}^nC_0 p^0 q^{n-0} = {}^6C_0 (0.554)^6 = (0.554)^6 = 0.02891$

$P(x=1) = {}^6C_1 (0.554)^5 (0.446)^1 = 6(0.446)(0.554)^5 = 0.1396$

$P(x=2) = {}^6C_2 (0.446)^2 (0.554)^4 = \frac{\cancel{6}^3\times5}{\cancel{2}}(0.446)^2(0.554)^4 = 0.2809$

$P(x=3) = {}^6C_3 (0.446)^3 (0.554)^3 = \frac{\cancel{6}\times5\times4}{\cancel{3\times2}}(0.446)^3(0.554)^3 = 0.3016$

$P(x=4) = {}^6C_4 (0.446)^4 (0.554)^2 = \frac{6\times5}{2}(0.446)^4(0.554)^2 = 0.1821$

$P(x=5) = {}^6C_5 (0.446)^5 (0.554)^1 = 6(0.446)^5(0.554) = 0.05864$

$P(x=6) = {}^6C_6 (0.446)^6 (0.554)^0 = (0.446)^6 = 0.007866$

$\therefore$ B·D is

x	0	1	2	3	4	5	6
f	13	25	52	58	32	16	4
E(x)	6	28	56	60	36	12	2

where $E(x)$ values are rounded off to nearest integer.

(11) (Pg 147 Pb45) Find the probability of getting an even number 3 or 4 or 5 times in throwing 10 dice, using B·D.

Sol: Let p = prob. of getting even no. We have 2, 4, 6 as possible even no's on a dice $\therefore p = \frac{3}{6} = \frac{1}{2} \Rightarrow q = \frac{1}{2}$.

Given $n = 10$. To find $P(x=3) = ?$ $P(x=4) = ?$; $P(x=5) = ?$

$\therefore P(x=3) = {}^{10}C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^7 = \frac{10\times9\times8}{3\times2\times1} \times \frac{1}{2^{10}} = \frac{120}{1024} = 0.112$

$P(x=4) = {}^{10}C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^6 = \frac{10\times9\times8\times7}{4\times3\times2\times1}\left(\frac{1}{2^{10}}\right) = \frac{210}{1024} = 0.2$

$P(x=5) = {}^{10}C_5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^5 = \frac{10\times9\times8\times7\times6}{5\times4\times3\times2}\left(\frac{1}{2^{10}}\right) = \frac{252}{1024} = 0.246$

$\left.\right\}$ .558

(6) Find the probability of getting an even number 3 or 4 or 5 times ~~with~~ in throwing 10 dice, using B.D.

Sol: Given $n = 10$

we know no. of even no.s on dice $= 3$ ($\because$ 2, 4, 6 are the no.s)

$\therefore p =$ prob. of getting an even number $= \frac{3}{6} = \frac{1}{2}$

$\Rightarrow q = 1 - p = \frac{1}{2}$.

$P(X=3) + P(X=4) + P(X=5) = {}^{10}C_3 p^3 q^7 + {}^{10}C_4 p^4 q^6 + {}^{10}C_5 p^5 q^5$

$= \frac{10\times 9\times 8}{3\times 2}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^7 + \frac{10\times 9\times 8\times 7}{4\times 3\times 2}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^6 + \frac{10\times 9\times 8\times 7\times 6}{5\times 4\times 3\times 2}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5$

$= \left(\frac{1}{2}\right)^{10}[120 + 210 + 252] = 0.112 + 0.2 + 0.246$

$= \frac{582}{1024} =$

(7) A coin is biased 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.

Sol: $P(H) = 2P(T)$; Given $n = 3$, To find $P(X = 2\text{ Tail } \& \text{ 1 Head}) = ?$

w.k.t $P(H) + P(T) = 1$ $\Rightarrow 2P(T) + P(T) = 1 \Rightarrow P(T) = \frac{1}{3}$

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

$P(x=2) = ? = {}^3C_2\, p^2 q^1 = \frac{3\times 2}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right) = \frac{2}{9}$ Ans.

(8) Out of 800 families with 5 children each, how many could you expect to have (a) 3 boys (b) 5 girls (c) either 2 or 3 boys (d) atleast 1 boy? Assume equal probabilities for boys & girls.

Sol: Given $N = 800$, $n = 5$. Let p - prob. of getting boy is success.

$\therefore p = \frac{1}{2}$; $q = \frac{1}{2}$ {$\because$ equal probabilities for boys & girls}

(a) $P(X=3) = {}^5C_3\, p^3 q^2 = \frac{5\times 4}{2}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{5}{16}$ per family.

$\therefore$ Total no. of families with 3 boys $= N \cdot P(x) = \frac{800}{16}\times 5 = 250$ families

(b) $P(X=0)$ ($\because$ all are girls ie 5 girls) $= {}^5C_0\, p^0 q^5 = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$

$\therefore$ Total no. of families with 5 girls $= 800\times\frac{1}{32} = 25$ families.

(c) $P(x=2)+P(x=3) = {}^5C_2 p^2 q^3 + {}^5C_3 p^3 q^2 = \frac{5\times \cancel{4}^{2}}{\cancel{2}}\left(\frac{1}{2}\right)^5 + \frac{5\times \cancel{4}^{2}}{\cancel{2}}\left(\frac{1}{2}\right)^5 = \frac{\cancel{20}^{5}}{\cancel{32}_{8}} = \frac{5}{8}$

$\therefore$ total no. of families with either 2 or 3 boys $= \cancel{800}^{100} \times \frac{5}{\cancel{8}} = 500$ families.

(d)

Problems on Poisson Distribution.

(i) (Pg 159 Pb 5) If a bank received on the average 6 bad cheques per day, find the probability that it will receive 4 bad cheques on any day Sol, $\lambda=6$ $P(x=4)=0.1339$..

Sol. $P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$; Given $\lambda=6$. To find $P(x=4)=?$

$\therefore P(x=4) = \frac{e^{-6}6^4}{4!} = \frac{54}{e^6} = 0.1339$.

(2) A car-hire firm has 2 cars which it hires out day by day. The no. of demands for a car on each day is distributed as a Poisson distribution with mean 1.5. Calculate the proportion of day (i) on which there is no demand (ii) on which demand is refused. (Pg 158 Pb 2)

Sol. Given $\lambda=1.5$; Total cars $=2$; $N=365$ days in a year

No demand $\Rightarrow$ cars were not hired $\Rightarrow P(x=0)=?$

No. of days when there is no demand in a year $= E(X=0)=?$ $= N.P(X=0)$.

(i) $P(x=0) = \frac{e^{-1.5}(1.5)^0}{0!} = e^{-1.5} = 0.2231$.

$\therefore$ No. of days in a year when there is no demand of car $= N \cdot P(X=0)$

$= 365 \times 0.2231$

$= 81$ days.

(ii) First we find probability on which demand is refused. w.r.t refusal takes place only when demand is for more than 2 cars. ie To find $P(X>2)$

To calculate proportion of days when there is refusal of demand $= E(X>2)=?$

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

$=1-e^{-1.5}\left[0+(1.5)+\frac{(1.5)^2}{2}\right] = \underline{0.1913}$

$E(X>2) = NP(X>2) = 365(0.1913) = 69.82 \simeq \underline{70 \text{ days}}$

(3) (Pg 160 PbP) It has been found that 2% of the tools produced by a certain machine are defective. What is the probability that in a shipment of 400 such tools (a) 3% or more (b) 2% or ~~more~~ less will prove defective.

Sol: No. of tools = 400.

% of defective tools = 2% = P

λ = No. of defective tools for 400 = nP = (2%) $\times$ (400)

$= 400 \times \frac{2}{100}$

$\Rightarrow \boxed{\lambda = 8}$

(i) $P(x \geq 3\%) = ?$

$3\% \Rightarrow \frac{3}{100} \times 400 = 12$ tools

$\therefore P(x \geq 3\%) = P(x \geq 12) = 1 - P(x \leq 11)$

$= 1 - e^{-\lambda}\left[\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \cdots + \frac{\lambda^{11}}{11!}\right]$

$= 1 - e^{-8}\left[1 + 8 + \frac{8^2}{2} + \frac{8^3}{3!} + \cdots + \frac{8^{11}}{11!}\right]$

$= 1 - e^{-8}\,[2647.29)$

$= 0.1119.$

(ii) $P(x \leq 2\%)$

$2\% \Rightarrow \frac{2}{100} \times 400 = 8$ tools

$\therefore P(x \leq 2\%) = P(x \leq 8) = P(x=0) + P(x=1) + P(x=2) + P(x=3) + P(x=4) + P(x=5) + P(x=6) + P(x=7) + P(x=8)$

$= e^{-\lambda}\left[\lambda^0 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \frac{\lambda^4}{4!} + \frac{\lambda^5}{5!} + \frac{\lambda^6}{6!} + \frac{\lambda^7}{7!} + \frac{\lambda^8}{8!}\right]$

$= e^{-8}\left[1 + 8 + \frac{\cancel{64}^{32}}{\cancel{2}} + \frac{8^3}{3!} + \frac{8^4}{4!} + \frac{8^5}{5!} + \frac{8^6}{6!} + \frac{8^7}{7!} + \frac{8^8}{8!}\right]$ $\frac{16777216}{40320}$

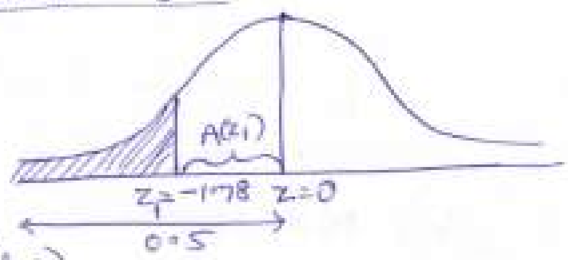
$= e^{-8}\,[9 + 32 + 85.3333 + 170.6667 + 273.0667 + 364.0889 + 416.1016 + 416.1016]$

$= \cancel{e^{-8}[1756.4238]} = \underline{0.5925}$

Problems related to Normal Distribution.

1. (Pg 205 pb10) If X is a normal variate, find the area A (i) to the left of $z = -1.78$ (ii) to the right of $z = -1.45$ (iii) corresponding to $-0.8 \le z \le 1.53$ (iv) to the left of $z = -2.52$ and to the right of $z = 1.83$.

Sol: (i) Required Area 'A' is 'shaded region'



Area $= 0.5 - A(z_1)$

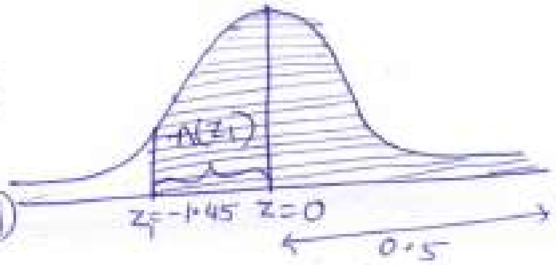
$= 0.5 - A(-1.78)$

$= 0.5 - A(1.78)$

$= 0.5 - 0.4625$ (from tables)

$= 0.0375$

(ii) Required Area $= 0.5 + A(z_1)$



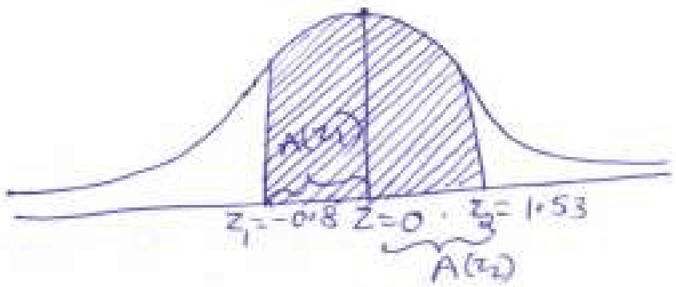
$= 0.5 + A(-1.45)$

$= 0.5 + A(1.45)$ (By symmetry)

$= 0.5 + 0.4265$

$= 0.9265$

(iii) Area $= A(z_1) + A(z_2)$



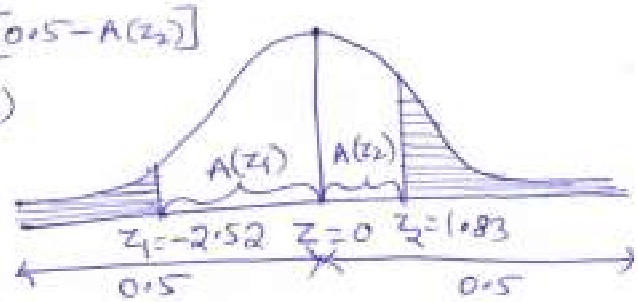
$= A(-0.8) + A(1.53)$

$= A(0.8) + A(1.53)$

$= 0.2881 + 0.437$

$= 0.7251$

(iv) Required Area $= [0.5 - A(z_1)] + [0.5 - A(z_2)]$



$= 0.5 - A(-2.52) + 0.5 - A(1.83)$

$= 1 - A(2.52) - A(1.83)$

$= 1 - 0.4941 - 0.4664$

$= 1 - 0.9605$

$= 0.0395$

2. (Pg 208 Pb13) If the masses of 300 students are normally distributed with mean 68 kgs & standard deviation 3 kgs, how many students have masses

(i) Greater than 72 kg

(ii) Less than or equal to 64 kg

(iii) Between 65 & 71 kg inclusive.

Sol. No. of students = 300 = N.

$\mu = 68$; $\sigma = 3$

(i) No. of students with mass greater than 72 kg = $N \cdot P(X)$ when $x = 72$.

$$Z = \frac{x-\mu}{\sigma} \Rightarrow z = \frac{72-68}{3} = \frac{4}{3} = 1.33$$

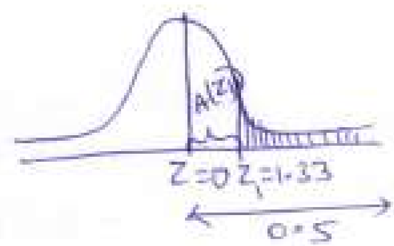
$\therefore P(X > 72) = P(Z > 1.33)$

$= 0.5 - A(z_1)$

$= 0.5 - A(1.33)$

$= 0.5 - 0.4082$

$= 0.0918$



$\therefore$ No. of students $= 300 \times 0.0918 = \boxed{28}$ students.

(ii) Less than or equal to 64 kg. $P(X \leq 64) = ?$

$$Z = \frac{x-\mu}{\sigma} = \frac{64-68}{3} = -\frac{4}{3} = -1.33$$

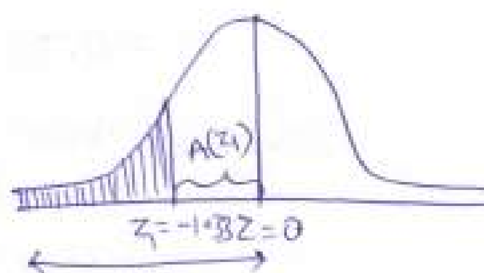
$\therefore P(X \leq 64) = P(Z \leq -1.33)$

$= 0.5 - A(z_1)$

$= 0.5 - A(-1.33)$

$= 0.5 - A(1.33)$

$= 0.5 - 0.4082 = \underline{0.0918}$



$\therefore$ No. of students $= 300 \times 0.0918 = \boxed{28}$ students.

(iii) $P(65 \leq X \leq 71)$ $\quad z_1 = \frac{65-68}{3} = -\frac{3}{3} = -1$; $z_2 = \frac{71-68}{3} = -1$

$= P(-1 \leq Z \leq 1)$

$= A(-1) + A(1)$

$= 2A(1)$

$= 2(0.3413)$

$= 0.6828$

$\therefore$ No. of students

$= 300 \times 0.6828$

$= \boxed{205}$ students

3 . (Pg 213 Pb 20) If x is normally distributed with mean 2 & S.D 0.1, then find $P(|x-2| \geq 0.01)$?

Sol: $\mu = 2$, $\sigma = 0.1$

$P(|x-2| \geq 0.01) = 1 - P(|x-2| < 0.01)$

$|x-2| < 0.01 \Rightarrow x = 2+0.01$ & $x = +2 - 0.01$

$= 2.01$ & $+1.99$

Let $x_1 = 1.99$, $x_2 = 2.01$

$z = \frac{x-\mu}{\sigma} \Rightarrow z_1 = \frac{1.99-2}{0.1} = \frac{-0.01}{0.1} = -0.1$

Also $z_2 = \frac{2.01-2}{0.1} = \frac{0.01}{0.1} = 0.1$

$\therefore P(|x-2| \geq 0.01) = 1 - P(z_1 < z < z_2)$

$= 1 - P(-0.1 < z < 0.1)$

$= 1 - [A(z_1) + A(z_2)]$

$= 1 - 2A(0.1)$

$= 1 - 2(0.0398)$

$= 1 - 0.0796 = \underline{0.9204}$

4. (Pg 222 Pb 20) Suppose the weights of 800 male students are normally distributed with mean 28.8 kg & S.D 2.06 kg. Find the no. of students whose weights are b/w (i) 28.4 kg & 30.4 kg (ii) more than 31.3 kg.

$N = 800$, mean $\mu = 28.8$; $\sigma = 2.06$

(i) $P(28.4 \leq x \leq 30.4) = ?$

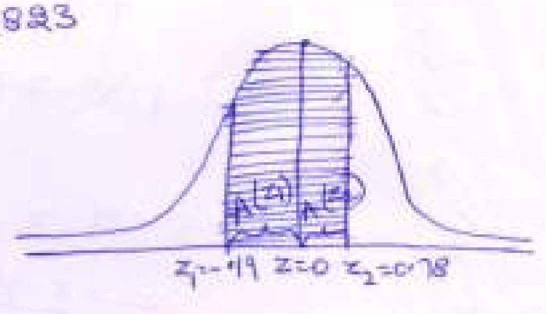
No. of students $= N \times P(28.4 \leq x \leq 30.4)$; $z = \frac{x-\mu}{\sigma}$

$x_1 = 28.4$, $x_2 = 30.4$

$z_1 = \frac{28.4 - 28.8}{2.06} = -\frac{\cancel{0.4}\ 0.2}{\cancel{2.06}\ 1.03} = \frac{-20}{103} = -0.1942 = 0$ $= -0.19$

$z_2 = \frac{30.4 - 28.8}{2.06} = \frac{\cancel{1.6}\ 0.8}{\cancel{2.06}\ 1.03} = \frac{80}{103} = 0.7767$ $= 0.78$

$$P(28.4 \le X \le 30.4) = P(z_1 \le Z \le z_2)$$
$$= P(-0.19 \le Z \le 0.78)$$
$$= A(z_1) + A(z_2)$$
$$= A(-0.19) + A(0.78)$$
$$= 0.0753 + 0.2823$$
$$= 0.3576$$

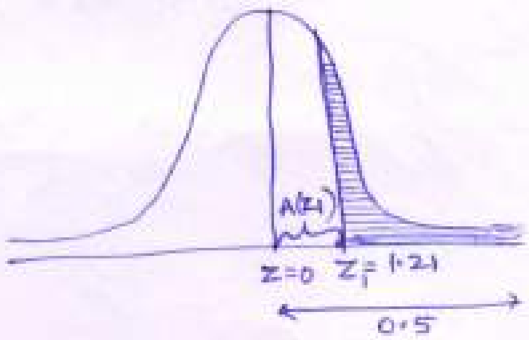


$\therefore$ No. of students $= N\,P(28.4 \le x \le 30.4)$
$= 800 \times 0.3576$
$= 286$

(ii) more than 31.3kg.

$P(X \ge 31.3)$

$$z = \frac{31.3 - 28.8}{2.06} = \frac{2.5}{2.06} = 1.21$$



$$P(X \ge 31.3) = P(Z \ge 1.21)$$
$$= 0.5 - A(z_1)$$
$$= 0.5 - A(1.21)$$
$$= 0.5 - 0.3869$$
$$= 0.1131$$

$\therefore$ No. of students $= N\,P(x \ge 31.3) = 800 \times 0.1131 = 90$

(Pg 200 Pb3)

✗ 5 In a Normal distribution, 7% of the items are under 35 and 89% are under 63. Determine mean & variance of the distribution.

$P(X < 35) = 7\%$. $P(X < 63) = 89\%$. $\mu = ?$, $\sigma^2 = ?$

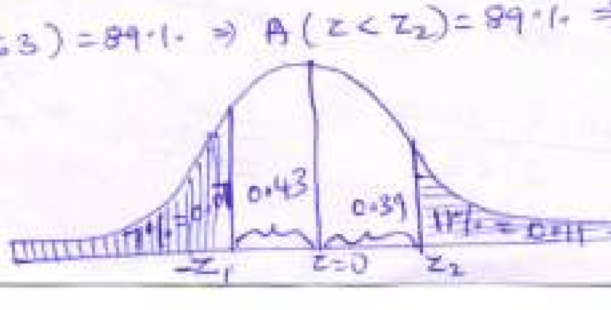
$x_1 = 35$, $x_2 = 63$ $\Rightarrow z_1 = \frac{35-\mu}{\sigma}$; $z_2 = \frac{63-\mu}{\sigma}$.

$\because$ Total $= 100\{\%\}$ we consider values of $x < 50$ to be left of $z=0$ {ie $x=\mu$} & values of $x > 50$ to be rt. of $z=0$. $\Rightarrow z_1 < 0$ & $z_2 > 0$ $\Rightarrow \frac{35-\mu}{\sigma} = -z_1$ (say)

$\frac{63-\mu}{\sigma} = z_2$ (say). $P(X < 35) = 7\% \Rightarrow A(Z < z_1) = 7\% \Rightarrow A(z_1) = 0.5 - 0.07 = 0.43$

$P(X < 63) = 89\% \Rightarrow A(Z < z_2) = 89\% \Rightarrow A(Z \ge z_2) = 11\%$. The fig is

$\Rightarrow A(z_2) = 0.5 - 0.11$
$= .39$



Next page.

Given $P(x<35) = 7\% = 0.07$; $P(x<63) = 89\% = 0.89$

when $x = 35$, $z = \frac{x-\mu}{\sigma} = \frac{35-\mu}{\sigma} = -z_1$ say ——(1)

when $x = 63$, $z = \frac{x-\mu}{\sigma} = \frac{63-\mu}{\sigma} = z_2$ say ——(2)

From the fig: $P(0<z<z_2) = 0.39 \Rightarrow A(z_2) = 0.39$

& $P(0<z<z_1) = 0.43 \Rightarrow A(z_1) = 0.43$.

Using tables we find $z_1 = 1.48$, $z_2 = 1.23$.

$\Rightarrow z_1 = -1.48$, $z_2 = 1.23$.

$\therefore$ (1) & (2) are $\frac{35-\mu}{\sigma} = -1.48 \Rightarrow 35-\mu = -1.48\sigma$.

$\Rightarrow \mu - 1.48\sigma = 35$ ——(3)

$\frac{63-\mu}{\sigma} = 1.23 \Rightarrow 63 = \mu + 1.23\sigma \Rightarrow \mu + 1.23\sigma = 63$ ——(4)

Solving (3) & (4) $\sigma = \frac{28}{2.71} = 10.332$

$\Rightarrow \boxed{\sigma^2 = 106.75}$

From (3) $35 = \mu - 1.48(10.332) \Rightarrow \mu = 35 + 15.3 = 50.3$

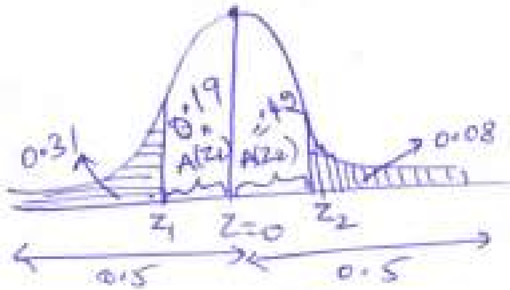
$\boxed{\therefore \mu = 50.3}$

*(6) In a N.D 31% of the items are under 45 & 8% are over 64. Find the mean & variance of the distribution. (Pg 202 P64)

$P(x<45) = 31\% = 0.31$; $P(X>64) = 8\% = 0.08$

$z = \frac{x-\mu}{\sigma}$ $\Rightarrow \underset{(\text{say})}{-z_1} = \frac{45-\mu}{\sigma}$; $z_2 = \frac{64-\mu}{\sigma}$.

The fig.:



From table

Area = 0.19 when $z = 0.5$

$\Rightarrow \boxed{z_1 = -0.5}$

Area = 0.42 when $z = 1.4 \Rightarrow$

$\Rightarrow \boxed{z_2 = 1.4}$

$\Rightarrow \frac{45-\mu}{\sigma} = -0.5 \Rightarrow 45-\mu = -0.5\sigma \Rightarrow \mu - 0.5\sigma = 45$ — (1)

$\frac{64-\mu}{\sigma} = 1.4 \Rightarrow 64 = \mu + 1.4\sigma$ — (2)

from (1) & (2) we get $\sigma = \frac{19}{1.9} = 10 \Rightarrow \boxed{\sigma = 10}$

From (1) : $\mu = 45 + 0.5\sigma = 45 + 0.5(10) = 50 \Rightarrow \boxed{\mu = 50}$

(7) (Pg 203 P67)
The marks obtained in mathematics by 1000 students is normally distributed with mean 78% & S.D 11%. Determine

(i) How many students got marks above 90%

(ii) What was the highest mark obtained by the lowest 10% of the students.

(iii) Within what limits did the middle of 90% of the students lie.

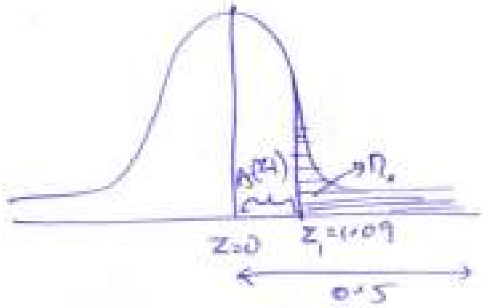
Sol: No. of students = 1000 ; $\mu = 78\% = 0.78$; $\sigma = 11\% = 0.11$

(i) $P(X > 90\%) = ?$ $\quad z = \frac{x-\mu}{\sigma} = \frac{0.9-0.78}{0.11} = 1.09 = Z_1$

$P(X>90\%) = P(Z>Z_1) = P(Z>1.09) = 0.5 - A(Z_1=1.09)$

$= 0.5 - 0.3621$

$= .1379$

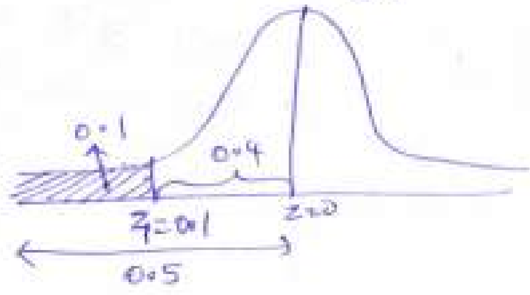


$\therefore$ No. of Students who got marks above 90% $= 1000\, P(X>90\%)$

$= 1000 \times 0.1379$

$= 137.9$

$= 138$ students.



(ii) $P(X<X_1) = 10\% = 0.1$

$\Rightarrow P(Z<Z_1) = 0.1$.

$\Rightarrow A(Z_1) + A(Z<Z_1) = 0.5$

$\Rightarrow A(Z_1) = 0.5 - 0.1 = 0.4$

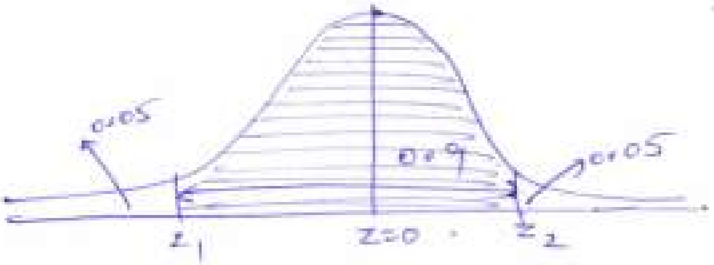
From Tables $Z_1 = 1.28 \Rightarrow Z_1 = -1.28$ {$\because$ left side of z}.

$-1.28 = \frac{x-\mu}{\sigma} = \frac{x-0.78}{0.11} \Rightarrow \underline{x = 0.6392}$

Hence the highest mark obtained by the lowest 10% of students $= 0.6392 \times 100 = 64\%$.

Next page.

7(iii)



To find $z_1 = ?$ $z_2 = ?$ such that $A(z_1) + A(z_2) = 0.9$

Middle 90% $\Rightarrow$ Area in the middle $= 0.9$.

$\Rightarrow$ Area on both sides left over $= 0.05$.

From table : value of z whose area $= 0.45$ is $z = 1.64$

$\because$ Middle area $= 0.9$ $\Rightarrow$ $A(z_1) = A(z_2) = \frac{0.9}{2} = 0.45$

From table Area $= 0.45$ for $z = 1.64$.

$\therefore z_1 = -1.64$; $z_2 = 1.64$.

$\Rightarrow \frac{x_1 - \mu}{\sigma} = -1.64$; $\frac{x_2 - \mu}{\sigma} = 1.64$.

$\Rightarrow x_1 = \mu - 1.64\sigma$; $x_2 = \mu + 1.64\sigma$.

Given $\mu = .78$; $\sigma = 0.11$.

$\therefore x_1 = 0.5996$; $x_2 = 0.9604$.

$\therefore$ Limits of 90% are $0.5996 \times 100 = 59.96 = 60\%$

& $0.9604 \times 100 = 96.04 = 96\%$.

$\Rightarrow$ B/W 60 & 96 ~~students~~ marks.

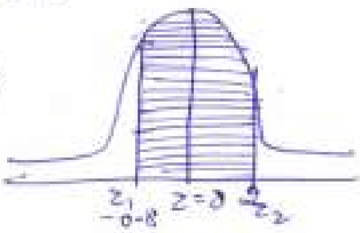
8. If X is a normal variate with mean 30 & standard deviation 5, find the probabilities that (i) $26 \le X \le 40$ (ii) $X \ge 45$.

Sol: $\mu = 30$, $\sigma = 5$; $z = \frac{x - \mu}{\sigma}$

(i) $P(26 \le X \le 40) = ?$

$x_1 = 26 \rightarrow z_1 = \frac{26 - 30}{5} = \frac{-4}{5} = -0.8$

$x_2 = 40 \Rightarrow z_2 = \frac{40 - 30}{5} = \frac{10}{5} = 2$



Thus $P(26 \le X \le 40) = P(-0.8 \le Z \le 2)$

$= A(0.8) + A(2)$

$= 0.2881 + 0.4772$

$= 0.7653$ (using Normal distribution tables)

Problems on Joint Probability [ie Multiple Random Variables]

(1) Pg 237 Pb 2. For following bivariate p.d. of X & Y find

(i) $P(X \le 2, Y = 2)$ (ii) $P(X < 3, Y \le 4)$ (iii) $P(Y = 3)$

(iv) $F_X(2)$ (v) $F_Y(3)$.

X/Y	1	2	3	4
1	0.1	0	0.2	0.1
2	0.05	0.12	0.08	0.01
3	0.1	0.05	0.1	0.09

(i) $P(X \le 2, Y = 2) = P(X = 1, Y = 2) + P(X = 2, Y = 2)$

$= 0 + 0.12$

$= 0.12$

(ii) $P(X < 3, Y \le 4) = P(X = 1, Y = 1) + P(X = 1, Y = 2) + P(X = 1, Y = 3)$
$+ P(X = 1, Y = 4) + P(X = 2, Y = 1) + P(X = 2, Y = 2)$
$+ P(X = 2, Y = 3) + P(X = 2, Y = 4)$

$= 0.1 + 0 + 0.2 + 0.1 + 0.05 + 0.12 + 0.08 + 0.01$

$= 0.66$

(iii) $P(Y = 3) = P(X = 1, Y = 3) + P(X = 2, Y = 3) + P(X = 3, Y = 3)$

$= 0.2 + 0.08 + 0.1$

$= 0.38.$

(iv) $F_X(2) = P(X = 2, Y = 1) + P(X = 2, Y = 2) + P(X = 2, Y = 3) + P(X = 2, Y = 4)$
$+ P(X = 1, Y = 1) + P(X = 1, Y = 2) + P(X = 1, Y = 3) + P(X = 1, Y = 4)$

$\{\because F_X(2) = P(X \le 2)\}$ $= 0.05 + 0.12 + 0.08 + 0.01 + 0.1 + 0 + 0.2 + 0.1$

$= 0.66$

(v) $F_Y(3) = P(Y \le 3) = P(X = 1, Y = 1) + P(X = 1, Y = 2) + P(X = 1, Y = 3)$
$+ P(X = 2, Y = 1) + P(X = 2, Y = 2) + P(X = 2, Y = 3)$
$+ P(X = 3, Y = 1) + P(X = 3, Y = 2) + P(X = 3, Y = 3)$

$= 0.1 + 0 + 0.2 + 0.05 + 0.12 + 0.08 + 0.1 + 0.05 + 0.1$

$= 0.70$

Unit II

Correlation

INTRODUCTION

In a bivariate distribution and multivariate distributions we may be interested to find if there is any relationship between the two variables under study.

The Correlation is a statistical tool which studies the relationship between two variables and Correlation analysis involves various methods and techniques used for studying and measuring the extent of the relationship between them.

→ Two variables are said to be Correlated if the change in one variable results in a corresponding change in the other variable.

Types of Correlation

1) Positive and Negative Correlation

If the values of the two variables deviate in the same direction i.e If the increase in the values of one variable results, on an average, in a corresponding increase in the value of the other variable (or) If the decrease in the values of one variable results, on an average, in a corresponding decrease in the value of the other variable, Correlation is said to be positive Correlation.

Eg:- 1. Heights and weights of the individuals.
2. The family income and expenditure on luxury items
3. Amount of Rainfall and yield of the Crop.

→ If the increase (decrease) in one variable results on an average, in a corresponding decrease (increase) in the values of the other variable, correlation is said to be <u>Negative Correlation</u>

Eg: 1. Price and demand of a commodity.
2. Sale of wollen garments and the day temperature
3. Volume and pressure of a perfect gas.

2. <u>Linear</u> and <u>Nonlinear</u> <u>Correlation</u>

→ The correlation between two variables is said to be <u>linear</u> if corresponding to a unit change in one variable, there is a constant change in the other variable over the entire range of the values. (or)

Two variables x and y are said to be linearly related if there exists a relationship of the form $y = a + bx$.

→ Two variables are said to be non linear or curvelinear if corresponding to a unit change in one variable, the other variable doesnot change at a constant rate but at fluctuating rate.

Methods of Studying Correlation

1) Scatter diagram method.
2) Karl pearson's coefficient of Correlation.
3) Rank method.

Scatter Diagram Method

If 'n' pair of values for two variables x and y are given as $(x_1, y_1), (x_2, y_2), \cdots (x_n, y_n)$. Then these points are dotted on the x-axis and y-axis in the xy plane. The diagram of dots so obtained is known as scatter diagram.

Note:- This method gives a fairly good, though rough, idea about the relationship between the two variables.

→ If the points are very dense (close to each other), then we say a fairly good amount of correlation between two variables.

→ If the points are widely scattered, a poor correlation between those variables.

→ If there is an upward trend rising from ~~left~~ the lower left hand corner and going upward to the upper right hand corner, the correlation is positive

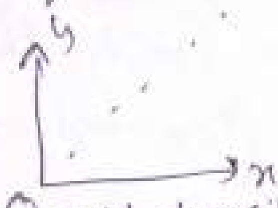
+ve Correlation

Negative Correlation

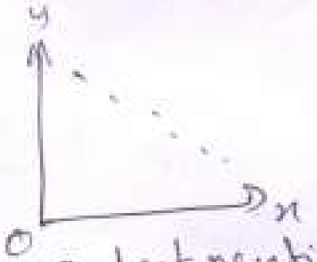
→ If the points depict a downward trend from the upper left hand corner to the lower right hand corner, the correlation is negative.

→ If all the points, lie on straight line from the left bottom and ~~getting~~ going up towards the right top, the correlation is said to be perfect positive

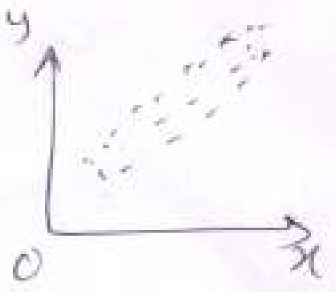
→ If all the points, lie on straight line from the left top and coming down toward right bottom, the correlation is said to be perfect Negative



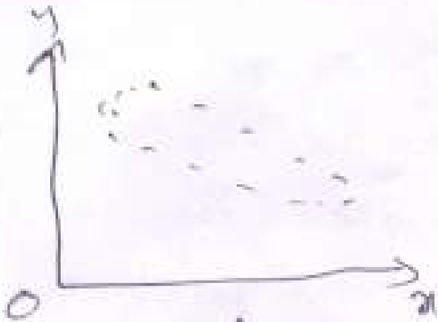
perfect positive



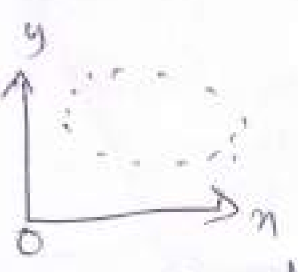
Perfect negative



High degree of +ve Correlation



High degree of -ve Correlation



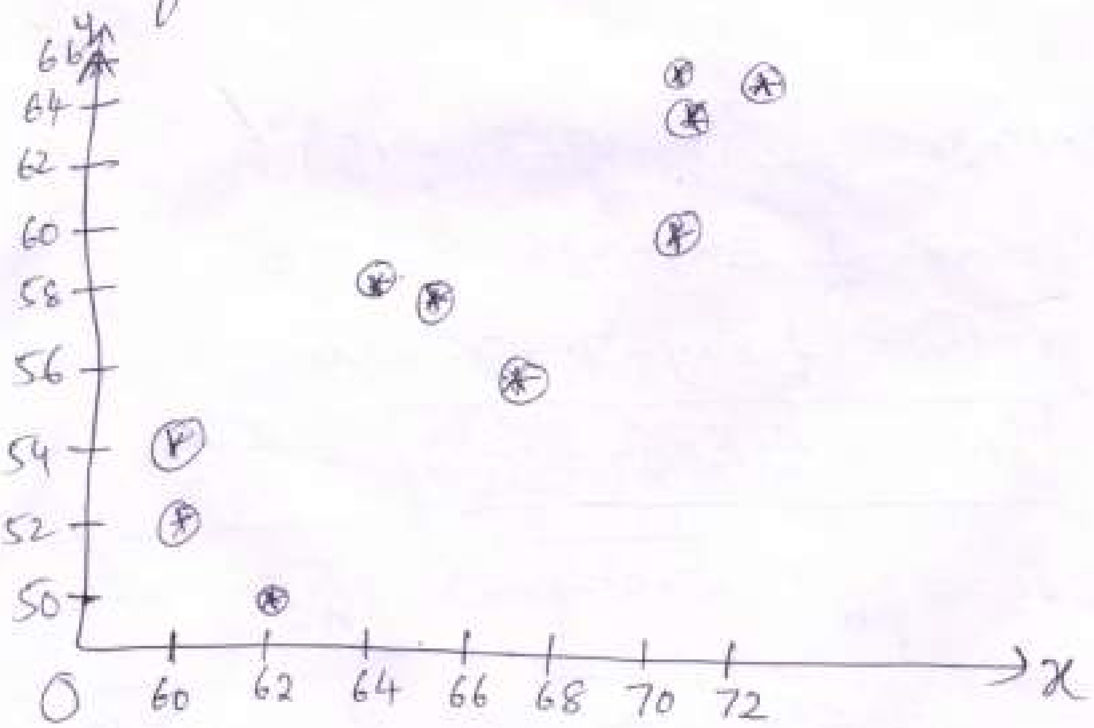
No Correlation

Prob1 Draw a scatter diagram from the following data

Height (inches) :	62	72	70	60	67	70	64	65	60	70
weight (lbs) :	50	65	63	52	56	60	59	58	54	65

Also indicate whether Correlation is positive or negative.

Sol.^{n} Let us take Heights as random variable 'x' and taking those values on X-axis. Let us take weights as variable 'y' and taking its values on Y-axis then we get scatter diagram as follows.



Since the points are close to each other we can say High degree of Correlation (positive) between the series of heights and weights.

Karl Pearson's Coefficient of Correlation.

Pearson's coefficient of correlation between two variables (series) X and Y is denoted by 'r' is a measure of linear relationship between them and is defined as

$$r = \frac{Cov(x,y)}{\sigma_x \sigma_y} \longrightarrow (1)$$

where $Cov(x,y) = \frac{1}{n}\sum(x-\bar{x})(y-\bar{y})$

$$\sigma_x = \sqrt{\frac{\sum(x-\bar{x})^2}{n}} \text{ and } \sigma_y = \sqrt{\frac{\sum(y-\bar{y})^2}{n}}$$

*Note 1:

$$Cov(x,y) = \frac{1}{n}\sum(xy - x\bar{y} - \bar{x}y + \bar{x}\bar{y})$$

$$Cov(x,y) = \frac{1}{n}\sum xy - \frac{\sum x}{n}\bar{y} - \bar{x}\frac{\sum y}{n} + \frac{\bar{x}\bar{y}}{n}\sum 1$$

$$Cov(x,y) = \frac{1}{n}\sum xy - \bar{x}\bar{y} - \cancel{\bar{x}\bar{y}} + \cancel{\bar{x}\bar{y}}$$

$$\boxed{Cov(x,y) = \frac{1}{n}\sum xy - \frac{\sum x}{n}\left(\frac{\sum y}{n}\right)} \longrightarrow (2)$$

Illy

$$\sigma_x = \sqrt{\frac{1}{n}\sum(x^2+\bar{x}^2-2x\bar{x})} = \sqrt{\frac{1}{n}\sum x^2 + \bar{x}^2 - 2\bar{x}^2}$$

$$\therefore \sigma_x = \sqrt{\frac{1}{n}\sum x^2 - \left(\frac{\sum x}{n}\right)^2} = \frac{1}{n}\sqrt{n\sum x^2 - (\sum x)^2}$$

$$\therefore \sigma_x = \frac{1}{n}\sqrt{n\sum x^2 - (\sum x)^2} \longrightarrow (3)$$

$$\text{and } \sigma_y = \frac{1}{n}\sqrt{n\sum y^2 - (\sum y)^2} \longrightarrow (4)$$

Sub (2), (3), (4) in (1) we get

$$\boxed{r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}} \longrightarrow (5)$$

Note 2:- $r = \dfrac{\Sigma dx\, dy}{\sqrt{\Sigma dx^2 \, \Sigma dy^2}} \rightarrow ⑥$ where $dx = x-\bar{x}$, $dy = y-\bar{y}$.

→ This formula is used when $\bar{x}$ and $\bar{y}$ are not fractionals (i.e not decimals or when natural numbers)

→ If the values of x and y are large, $\bar{x}, \bar{y}$ are fractionals then using this formula becomes tedious. So, we use Step deviation method

→ If the values of x and y are small but $\bar{x}, \bar{y}$ are fractionals then we use formula ⑤

Properties of Correlation Coefficient

** 1. The correlation coefficient lies between -1 and 1 i.e $-1 \le r \le 1$.

Proof:- We know that $\Sigma \left[\dfrac{x-\bar{x}}{\sigma_x} \pm \dfrac{y-\bar{y}}{\sigma_y} \right]^2 \ge 0$

$$\Sigma \left\{ \left[\frac{x-\bar{x}}{\sigma_x} \right]^2 + \left[\frac{y-\bar{y}}{\sigma_y} \right]^2 \pm \frac{2(x-\bar{x})(y-\bar{y})}{\sigma_x \sigma_y} \right\} \ge 0$$

$$\Rightarrow \frac{\Sigma (x-\bar{x})^2}{\sigma_x^2} + \frac{\Sigma (y-\bar{y})^2}{\sigma_y^2} \pm \frac{2}{\sigma_x \sigma_y} \Sigma (x-\bar{x})(y-\bar{y}) \ge 0$$

Dividing with 'n' on both sides, we get

$$\frac{\sigma_x^2}{\sigma_x^2} + \frac{\sigma_y^2}{\sigma_y^2} \pm \frac{2\,\mathrm{CoV}(x,y)}{\sigma_x \sigma_y} \ge 0 \quad \left[\because \frac{\Sigma (x-\bar{x})^2}{n} = \sigma_x^2 \right]$$

$$2 \pm 2r \ge 0$$

$2-2r \ge 0$	$2+2r \ge 0$
$2 \ge 2r \Rightarrow 1 \ge r$	$2 \ge -2r$
$\Rightarrow r \le 1$	$1 \ge -r$
	$-1 \le r$

$\boxed{\therefore -1 \le r \le 1}$

→ If $r = 1 \Rightarrow$ +ve correlation (perfect)
$r = -1 \Rightarrow$ −ve Correlation (perfect)
$r = 0 \Rightarrow$ Uncorrelated (No correlation)

2. The coefficient of Correlation is independent of the change of origin and scale.

<u>proof</u>:- Let x and y are transformed into new variables u and v by the change of origin and scale

i.e $\boxed{u = \frac{x-A}{h}}$, $\boxed{v = \frac{y-B}{K}}$

where $h, K > 0$ and A, B, h, k are constants

then $x = A + uh$ and $y = B + vK$

$$\left.\begin{aligned} \Sigma x &= \Sigma(A + uh) \\ \frac{\Sigma x}{n} &= \frac{nA}{n} + \frac{h\Sigma u}{n} \\ \Rightarrow \bar{x} &= A + \bar{u}h \end{aligned}\right\} \text{ slly we get } \bar{y} = B + \bar{v}K$$

Now $x - \bar{x} = h(u - \bar{u})$ and $y - \bar{y} = K(v - \bar{v})$

$$\therefore r_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y} = \frac{\Sigma hK(u-\bar{u})(v-\bar{v})}{n\sqrt{\frac{\Sigma(u-\bar{u})^2h^2}{n}\frac{\Sigma(v-\bar{v})^2K^2}{n}}}$$

$$\therefore r_{xy} = \frac{\Sigma(u-\bar{u})(v-\bar{v})}{\sqrt{\Sigma(u-\bar{u})^2\Sigma(v-\bar{v})^2}} = \frac{Cov(u,v)}{\sigma_u\sigma_v} = r_{uv}$$

$$\boxed{\therefore r_{xy} = r_{uv}}$$

<u>Note</u>:- $r = \frac{\Sigma du\,dv}{\sqrt{\Sigma du^2 \Sigma dv^2}} = \frac{n\Sigma uv - \Sigma u\Sigma v}{\sqrt{[n\Sigma u^2 - (\Sigma u)^2][n\Sigma v^2 - (\Sigma v)^2]}}$ is used when $\bar{x}$ or/and $\bar{y}$ are fractionals or If x, y values are large.

Problem 1. Calculate coefficient of correlation for the following data

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

Solution Given $n = 9$.

$\therefore \bar{x} = \frac{\Sigma x}{n} = \frac{45}{9} = 5$ and $\bar{y} = \frac{\Sigma y}{n} = \frac{108}{9} = 12$.

Since the given values are small and $\bar{x}, \bar{y}$ are not fractionals,

correlation coefficient is given by

$$r = \frac{\Sigma dx\, dy}{\sqrt{\Sigma dx^2 \Sigma dy^2}} \quad \text{where} \quad \begin{aligned} dx &= x - \bar{x} \\ dy &= y - \bar{y} \end{aligned}$$

Now the table is

x	y	$dx = x - \bar{x}$ ($x-5$)	$dy = y - \bar{y}$ ($y-12$)	dx^2	dy^2	$dx\,dy$
9	15	4	3	16	9	12
8	16	3	4	9	16	12
7	14	2	2	4	4	4
6	13	1	1	1	1	1
5	11	0	−1	0	1	0
4	12	−1	0	1	0	0
3	10	−2	−2	4	4	4
2	8	−3	−4	9	16	12
1	9	−4	−3	16	9	12
$\Sigma x = 45$	$\Sigma y = 108$			$\Sigma dx^2 = 60$	$\Sigma dy^2 = 60$	$\Sigma dx dy = 57$

$$\therefore r = \frac{57}{\sqrt{60(60)}} = \frac{57}{60} = 0.95$$

$\therefore$ There is a high degree of positive correlation between x and y series.

2. Find correlation coeff from

x	10	12	18	24	23	27
y	13	18	12	25	30	10

3. Find Karl pearson's coefficient of correlation from

wages	100	101	102	102	100	99	97	98	96	95
Cost of living	98	99	99	97	95	92	95	94	90	91

Sol. Given $n = 10$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{2(100)+101+2(102)+99+97+98+96+95}{10} = \frac{990}{10} = 99$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{950}{10} = 95$$

The table is

x	y	$dx = x-\bar{x}$ $x-99$	$dy = y-\bar{y}$ $y-95$	dx^2	dy^2	$dxdy$
100	98	1	3	1	9	3
101	99	2	4	4	16	8
102	99	3	4	9	16	12
102	97	3	2	9	4	6
100	95	1	0	1	0	0
99	92	0	-3	0	9	0
97	95	-2	0	4	0	0
98	94	-1	-1	1	1	1
96	90	-3	-5	9	25	15
95	91	-4	-4	16	16	16
990 Σx	950 Σy			$\Sigma dx^2 = 54$	$\Sigma dy^2 = 96$	$\Sigma dxdy = 61$

$$\therefore r = \frac{\Sigma dxdy}{\sqrt{\Sigma dx^2 \Sigma dy^2}} = \frac{61}{\sqrt{54(96)}} = \frac{61}{705.453} = \cancel{0.088}\ 0.847$$

$$\boxed{\therefore r = 0.847}$$

4. Calculate the coeff of correlation between age of cars and annual maintenance cost and comment

Age of Cars (yrs)	2	4	6	7	8	10	12
Annual Maintenance cost (Rs)	1600	1500	1800	1900	1700	2100	2000

Sol. Since 'y' values are large, we use step deviation method.

Now, $n = 7$, $\bar{x} = \frac{\Sigma x}{n} = \frac{49}{7} = 7$

and $\bar{y} = \frac{\Sigma y}{n} = \frac{12600}{7} = 1800$

let $h = 1$ and $K = 100$ and $A = \bar{x}$; $B = \bar{y}$

then $u = \frac{x - \bar{x}}{h} = x - 7$ & $v = \frac{y - 1800}{100}$

The table is

x	y	$u = x-7$	$v = \frac{y-1800}{100}$	uv	u^2	v^2
2	1600	-5	-2	10	25	4
4	1500	-3	-3	9	9	9
6	1800	-1	0	0	1	0
7	1900	0	1	0	0	1
8	1700	1	-1	-1	1	1
10	2100	3	3	9	9	9
12	2000	5	2	10	25	4
49	12600	Σu 0	Σv 0	37 Σuv	70 Σu^2	28 Σv^2

$$\therefore r = \frac{n\Sigma uv - \Sigma u \Sigma v}{\sqrt{[n\Sigma u^2 - (\Sigma u)^2][n\Sigma v^2 - (\Sigma v)^2]}} = \frac{7(37)}{\sqrt{7^2(70)(28)}} = \frac{37}{\sqrt{1960}}$$

$$= 0.8357$$

$$\simeq 0.836$$

5. Find out the coeff of correlation from

Height of father (Inches)	65	66	67	68	69	71	73
Height of Son (inches)	67	68	64	72	70	69	70

Sol. Given $n=7$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{65+66+67+68+69+71+73}{7} = 68.428$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{67+68+64+72+70+69+70}{7} = 68.57$$

Since $\bar{x}, \bar{y}$ are fractionals,

let $u = \frac{x-68}{1}$ and $v = \frac{y-69}{1}$. ($\because h=k=1$)

The table is

x	y	$u=x-68$	$v=y-69$	u^2	v^2	uv
65	67	-3	-2	9	4	6
66	68	-2	-1	4	1	2
67	64	-1	-5	1	25	5
68	72	0	3	0	9	0
69	70	1	1	1	1	1
71	69	3	0	9	0	0
73	70	5	1	25	1	5
		$\Sigma u=3$	$\Sigma v=-3$	$\Sigma u^2=49$	$\Sigma v^2=41$	$\Sigma uv=19$

$$\therefore r = \frac{n\Sigma uv - \Sigma u \Sigma v}{\sqrt{[n\Sigma u^2-(\Sigma u)^2][n\Sigma v^2-(\Sigma v)^2]}} = \frac{7(19)+9}{\sqrt{[7(49)-9][7(41)-9]}}$$

$$= \frac{142}{\sqrt{334(278)}} = \frac{142}{\sqrt{92852}} = \frac{142}{304.72} = 0.466$$

$$\boxed{\therefore r = 0.466}$$

7. Given $n = 10$, $\sigma_x = 5.4$, $\sigma_y = 6.2$ and sum of the product of deviation from the mean of X and Y is 66. find the correlation coeff.

Solⁿ Given $n = 10$, $\sigma_x = 5.4$, $\sigma_y = 6.2$ and $\Sigma(x-\bar{x})(y-\bar{y}) = 66$

$$\therefore \quad r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y} = \frac{66}{10(5.4)(6.2)} = \frac{6.6}{(5.4)(6.2)} = \underline{\underline{0.197}}$$

Rank Correlation Coefficient

Spearman's rank correlation is given by

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

where d_i = Difference of corresponding ranks of x and y.

n = no of the terms in the series

Properties

1. $-1 \leq \rho \leq 1$
2. If $\rho = 1$, there is complete agreement in the order of the ranks and they are in same direction.
3. If $\rho = -1$, there is complete disagreement in the order of the ranks and they are in opposite direction.

Problems when ranks are given

1. Following are the ranks obtained by 10 students in two subjects, Statistics and Maths. To what extent the knowledge of the students in two subjects is related?

Statistics	1	2	3	4	5	6	7	8	9	10
Maths	2	4	1	5	3	9	7	10	6	8

Sol. Given $n=10$; let x = Ranks in Statistics
y = Ranks in Maths

The table is

x	y	$d_i = x-y$	d_i^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
			$\Sigma d^2 = 40$

Rank Correlation Coefficient is given by

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

$$= 1 - \frac{6(\cancel{40})4}{\cancel{10}(10^2-1)}$$

$$= 1 - \frac{24}{99}$$

$$= \frac{75}{99} = 0.7575$$

$$\boxed{\therefore \rho = 0.76}$$

2. A random sample of 5 students are selected and their grades in Maths and statistics are found to be

	1	2	3	4	5
Maths	85	60	73	40	90
stats	93	75	65	50	80

Sol. Let x = Ranks in Maths.
y = Ranks in Stats $\quad n = 5$

[Here give ranks' based on their marks in decending order]

Marks in Maths	Rank x	Marks in Stats	Rank y	$d_i = x-y$	d_i^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
					4

$$\therefore \rho = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)} = 1 - \frac{6(4)}{5(5^2-1)} = 1 - \frac{\cancel{6(4)}}{5\cancel{(24)}}$$

$$\therefore \rho = 1 - \frac{1}{5} = \frac{4}{5} = 0.8$$

$$\boxed{\therefore \rho = 0.8}$$

Problems on equal or Repeated Ranks

Formula

$$\rho = 1 - 6\left\{\frac{\Sigma d^2 + \frac{1}{12}(m)(m^2-1) + \frac{1}{12}m(m^2-1) + \cdots}{n(n^2-1)}\right\}$$

Here $m \longrightarrow$ no of times rank is repeated.

Procedure

→ If the items are repeated in the series then common ranks are assigned such that the mean of the ranks which these items would have got if they were different from each other and the next item will get the rank next to the rank used in computing the common rank.

Eg: 1. An item is repeated at rank 4, twice then

Common rank = $\frac{4+5}{2} = \frac{9}{2} = 4.5$

and next rank will be '6'.

2. An item is repeated thrice at rank 3 then

Common Rank = $\frac{3+4+5}{3} = \frac{12}{3} = 4$

and next rank will be '6'.

Prob1. The ranks of the 15 students in two subjects A and B are given below, the two numbers within the brackets denoting the ranks of the same student in A and B resply.

(1,10), (2,7), (3,2), (4,6), (5,4), (6,8), (7,3), (8,1), (9,11), (10,15), (11,9)

(12,5), (13,14), (14,12) and (15,13).

Use Spearman's formula to find the rank correlation coefficient.

Sol:ⁿ [Since ranks not repeated use.

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

Ans: 0.514

where $n = 15$.

2. The following table gives the score obtained by 11 Students in English and Telugu. Find the rank correlation Coeff.

Scores of English	40	46	54	60	70	80	82	85	85	90	95
Scores of Telugu	45	45	50	42	48	75	55	72	65	42	70

Sol. Given $n=11$

Marks in English	Scores in Telugu	Rank x	Rank y	d_i	d_i^2
40	45	11	7.5	3.5	12.25
46	45	10	7.5 m=2	2.5	6.25
54	50	9	6	3	9
60	43	8	9	−1	1
70	40	7	11	−~~4~~	16
80	75	6	1	5	20
82	55	5	5	0	0
85	72	3.5	2	1.5	2.25
85	65	3.5 m=2	4	−0.4	0.16
90	42	2	10	−8	64
95	70	1	3	−2	4
					$\Sigma d_i^2 = 134.91$

∴ Spearman's rank correlation is given by

$$\rho = 1 - 6\left\{\frac{\Sigma d^2 + \frac{1}{12}m(m^2-1) + \frac{1}{12}m(m^2-1)}{n(n^2-1)}\right\}$$

$$= 1 - 6\left\{\frac{134.91 + \frac{1}{\cancel{6}}2(3)}{11(120)}\right\}$$

$$= 1 - \cancel{6}\left\{\frac{135.91}{\cancel{1320}\ 220}\right\} = 1 - 0.618 = 0.382$$

$\boxed{\therefore \rho = 0.382}$

2. A sample of 12 fathers and their elder sons gave the following data about their elder sons. Calculate the Coeff of rank correlation.

Fathers	65	63	67	64	68	62	70	66	68	67	69	71
Sons	68	66	68	65	69	66	68	65	71	67	68	70

Solⁿ Here $n = 12$. The table is

Fathers X	Sons Y	Rank x	Rank y	d_i	d_i^2
65	68	9	5.5	3.5	12.25
63	66	10	9.5	1.5	2.25
67	68	6.5	5.5	1	1
64	65	10	11.5	-1.5	2.25
68	69	4.5	3	1.5	2.25
62	66	12	9.5	2.5	6.25
70	68	2	5.5	-3.5	12.25
66	65	8	11.5	-3.5	12.25
68	71	4.5	1	3.5	12.25
67	67	6.5	8	-1.5	2.25
69	68	3	5.5	-2.5	6.25
71	70	1	2	-1	1
					72.5

In X series 68 repeated 2 times, 67 repeated two times

So, $m = 2, 2$.

In Y series 68 repeated 4 times, 65 repeated 2 times, 66 repeated 2 times

So, $m = 4, 2, 2$

$$\therefore \rho = 1 - 6\left\{\frac{\Sigma d^2 + \frac{1}{12}2(2^2-1)(2) + \frac{1}{12}4(4^2-1)}{12(12^2-1)}\right\} = 1 - 6\left\{\frac{72.5 + 2 + 5}{12(143)}\right\}$$

$$\therefore \rho = 1 - 6\left\{\frac{79.5}{1716}\right\} = 1 - \frac{79.5}{286} = 1 - 0.278 = 0.722$$

$$\boxed{\therefore \rho = 0.722}$$

Regression

The statistical method which helps us to estimate the unknown value of one variable from the known value of the related variable is called Regression.

The lines described in the average relationship between two variables is known as Lines of Regression

Comparision between Correlation and Regression

Correlation	Regression
1. It is a measure of degree of covariability between two variables.	1. Regression establishes the functional relationship between dependent & independent variables
2. In correlation, both the variables are random variables	2. In Regression, one variable is dependent variable and other one is independent variable
3. The coefficient of correlation is a relative measure	3. Regression Coefficient is an absolute measure

Linear Regression

→ If the Regression equation is a straight line then we say it is a Linear Regression. Otherwise we call it as Non-linear Regression

Lines of Regression

→ Line of Regression of 'y' on 'x' is the line which gives the best estimate for the value of y for any Specified Value of 'x'.

→ Line of Regression of 'x' on 'y' is the line which gives the best estimate for the value of x for any Specified value of 'y'.

Regression Equation of y on x $\Rightarrow$ $y = a + bx$ → (1)

$$na + b\Sigma x = \Sigma y$$
$$a\Sigma x + b\Sigma x^2 = \Sigma xy$$
(Normal equations)

Solving we get a, b values

Substituting in (1) we get Required Regression eq^n of y on x.

Regression Equation of x on y is $x = a + by$ → (2)

Normal equations are

$$na + b\Sigma y = \Sigma x$$
$$a\Sigma y + b\Sigma y^2 = \Sigma xy$$

Solving we get a, b values

Sub in (2) to get Required Regression Equation of x on y.

→ Regression Equations by using deviations from arithmetic mean of x and y.

Regression Equation of 'y' on 'x' is

$$y-\bar{y} = r\frac{\sigma_y}{\sigma_x}(x-\bar{x})$$

where $\bar{x}$ = Mean of 'x' series $= \frac{\Sigma x}{n}$.

$\bar{y}$ = Mean of 'y' series $= \frac{\Sigma y}{n}$

r = Correlation coefficient of x & y.

$\sigma_x, \sigma_y \rightarrow$ S.D of x, y series

Regression Coefficient of y on x $= r\frac{\sigma_y}{\sigma_x}$

$$= \frac{cov(x,y)}{\sigma_x \sigma_y} \cdot \frac{\sigma_y}{\sigma_x}$$

$$= \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\,\sigma_x^2}$$

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

$\Rightarrow$ Regression Coeff of y on x $= \frac{\Sigma XY}{\Sigma X^2}$

$\therefore$ Regression equation of 'y' on x is

$$\boxed{y-\bar{y} = \frac{\Sigma XY}{\Sigma X^2}(x-\bar{x})}$$

Slly Regression Equation of 'x' on y is

$$x-\bar{x} = r\frac{\sigma_x}{\sigma_y}(y-\bar{y})$$

$$\Rightarrow \boxed{x-\bar{x} = \frac{\Sigma XY}{\Sigma Y^2}(y-\bar{y})}$$

Here if Regression coeff of y on 'x' is b_{yx} and regression coeff of x on 'y' is b_{xy} then

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} \text{ and } b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

$$\Rightarrow b_{yx} \cdot b_{xy} = r^2$$

$$\Rightarrow \boxed{r^2 = b_{xy} \cdot b_{yx}}$$

<u>Note 1</u> Regression line always passes through $(\bar{x}, \bar{y})$

<u>Prob</u> If two regression lines of y on x and x on y are $a_1x + b_1y + c_1 = 0$ and $a_2x + b_2y + c_2 = 0$ then prove that $a_1b_2 < a_2b_1$

<u>Sol.</u> Given lines can also be written as
Regression eqn of y on x is

$$y = -\frac{a_1}{b_1}x - \frac{c_1}{b_1} = -\frac{c_1}{b_1} + \left(-\frac{a_1}{b_1}\right)x$$

$$\Rightarrow y = a + bx$$

Regression coeff of y on $x = b = \left(\frac{-a_1}{b_1}\right) = b_{yx}$

Sly we can find.

Regression coeff of x on $y = \left(\frac{-b_2}{a_2}\right) = b_{xy}$

But $r^2 = b_{yx} \cdot b_{xy} = \left(-\frac{a_1}{b_1}\right)\left(-\frac{b_2}{a_2}\right) = \frac{a_1b_2}{a_2b_1}$

Since correlation coeff $-1 \le r \le 1$

$$\Rightarrow \frac{a_1b_2}{a_2b_1} < 1 \Rightarrow \boxed{a_1b_2 < a_2b_1}$$

Angle between two Regression lines

We know that

Regression equation of y on x is

$$y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

$\Rightarrow$ Slope of line $= m_1 = r\dfrac{\sigma_y}{\sigma_x}$

Regression equation of x on y is

$$x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

$\Rightarrow$ Slope of line $= m_2 = \dfrac{1}{r}\dfrac{\sigma_y}{\sigma_x}$

$\therefore$ Angle between two lines $-$ $\tan\theta = \dfrac{m_1 - m_2}{1 + m_1 m_2}$

$$\therefore \tan\theta = \frac{r\frac{\sigma_y}{\sigma_x} - \frac{1}{r}\frac{\sigma_y}{\sigma_x}}{1 + \frac{\sigma_y^2}{\sigma_x^2}}$$

$$= \frac{\frac{\sigma_y}{\sigma_x}\left(r - \frac{1}{r}\right)}{\left(\frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^2}\right)}$$

$$\boxed{\tan\theta = \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\left(\frac{r^2 - 1}{r}\right)}$$

Notes: 1. If θ is acute, $\tan\theta = \dfrac{\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2}\left(\dfrac{1 - r^2}{r}\right)$

2. If θ is obtuse, $\tan\theta = \dfrac{\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2}\left(\dfrac{r^2 - 1}{r}\right)$

3. If $r = 0 \Rightarrow \tan\theta = \alpha \Rightarrow \boxed{\theta = \pi/2}$

$\Rightarrow$ No relation between two ~~regression~~ Variables

4. If $r = \pm 1 \Rightarrow$ the Variables $\theta = 0$ or π $\Rightarrow$ Two lines are Coincident or parallel

Pb1. Find the Regression line of x on y given

x	78	77	85	88	87	82	81	77	76	83	97	93
y	84	82	82	85	89	90	88	92	83	89	98	99

Sol^n Regression equation of x on y is

$$x-\bar{x} = r\frac{\sigma_x}{\sigma_y}(y-\bar{y})$$

$$\Rightarrow x-\bar{x} = \frac{\Sigma XY}{\Sigma Y^2}(y-\bar{y})$$

Here $\bar{x} = \frac{\Sigma x}{n} = \frac{78+77+\cdots+93}{12} = \frac{1004}{12} = 83.66$

$\bar{y} = \frac{\Sigma y}{n} = \frac{84+82+\cdots+99}{12} = \frac{1061}{12} = 88.41$

Since $\bar{x}, \bar{y}$ are not whole numbers, we use Assumed mean method,

$\therefore$ Regression eq^n of x on y is

$$x-\bar{x} = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma Y^2 - (\Sigma Y)^2}(y-\bar{y})$$

x	y	$X = x - A$ $= x - 84$	$Y = y - B$	Y^2	XY
78	84	−6	−4	16	24
77	82	−7	−6	36	42
85	82	1	−6	36	−6
88	85	4	−3	9	−12
87	89	3	1	1	3
82	90	−2	2	4	−4
81	88	−3	0	0	0
77	92	−7	4	16	−28
76	83	−8	−5	25	40
83	89	−1	+1	1	−1
97	98	13	10	100	130
93	99	9	11	121	99
1004	1061	−4	5	365	287

$\therefore$ Regression eq^n of x on y is

$$(x-83.66) = \frac{12(287)+20}{12(365)-25}(y-88.41) \Rightarrow \boxed{x = 0.795y + 13.38}$$

Problems on Regression

1. Find the most likely production corresponding to a rainfall 40 from the following data

	Rainfall X	Production Y
Average	30	500 kgs
S.D	5	100 kgs
r	0.8	

Sol^n We have to find Y value at $X = 40$.

So we have to find Regression eq^n of Y on X

Given $\bar{X} = 30$, $\sigma_x = 5$, $\sigma_y = 100$; $\bar{Y} = 500$; $r = 0.8$

Regression eq^n of Y on X is

$$(Y - \bar{Y}) = r\frac{\sigma_y}{\sigma_x}(X - \bar{X})$$

$$\Rightarrow Y - 500 = 0.8\left(\frac{100}{5}\right)(X - 30)$$

$$\Rightarrow Y - 500 = 8X - 240$$

$$\Rightarrow Y = 8X + 260$$

$\therefore$ Y at $X = 40 \Rightarrow Y(40) = 8(40) + 260$
$= 320 + 480$

$\therefore$ production corresponding to a rainfall $40 = 480$ kgs

2. From a sample of 200 pairs of the observation the following quantities were calculated.

$\Sigma X = 11.34$, $\Sigma Y = 20.78$, $\Sigma X^2 = 12.16$, $\Sigma Y^2 = 84.96$; $\Sigma XY = 22.13$

then find Regression eq^n of Y on X, compute the correlation coefficient.

Sol^n Regression eq^n of Y on X is $Y - \bar{Y} = \frac{\Sigma XY}{\Sigma X^2}(X - \bar{X})$

$$\Rightarrow \left(Y - \frac{20.78}{200}\right) = \frac{22.13}{12.16}\left(X - \frac{11.34}{200}\right)$$

Problems on Regression

1. Find the most likely production corresponding to a rainfall 40 from the following data.

	Rainfall X	Production Y
Average	30	500 kgs
S.D	5	100 kgs
r	0.8	

Sol^n we have to find Y value at $X = 40$.

So we have to find Regression eq^n of Y on X

Given $\bar{X} = 30$, $\sigma_x = 5$, $\sigma_y = 100$; $\bar{Y} = 500$; $r = 0.8$

Regression eq^n of Y on X is

$$(Y - \bar{Y}) = r\frac{\sigma_y}{\sigma_x}(X - \bar{X})$$

$$\Rightarrow Y - 500 = 0.8\left(\frac{\cancel{100}^{20}}{\cancel{5}}\right)(X - 30)$$

$$\Rightarrow Y - 500 = 8X - 240$$

$$\Rightarrow Y = 8X + 260$$

$$\therefore Y \text{ at } X = 40 \Rightarrow Y(40) = 8(40) + 260$$
$$= 320 + 480$$

$\therefore$ production corresponding to a rainfall 40 = 480 kgs.

2. From a sample of 200 pairs of the observation the following quantities were calculated.

$\Sigma X = 11.34$, $\Sigma Y = 20.78$, $\Sigma X^2 = 12.16$, $\Sigma Y^2 = 84.96$; $\Sigma XY = 22.13$

then find Regression eq^n of Y on X, compute the correlation coefficient.

Sol^n Regression eq^n of Y on X is $Y - \bar{Y} = \frac{\Sigma XY}{\Sigma X^2}(X - \bar{X})$

$$\Rightarrow \left(Y - \frac{20.78}{200}\right) = \frac{22.13}{12.16}\left(X - \frac{11.34}{200}\right)$$

The 8th performance for which Judge Q could not attend, was awarded 37 marks by judge P. If judge Q present, how many marks he will give to 8th perf.? (15)

3. Sol. Let 'x' be the series of Marks by P and

'y' be the Series of Marks by Q.

Here $\bar{x} = \frac{\Sigma x}{n} = \frac{46+42+44+40+43+41+45}{7} = 43.$

Slly $\bar{y} = \frac{\Sigma y}{n} = 38.$

Since $\bar{x}, \bar{y}$ are whole numbers, we use the Regression eqn of y on x as

$$y - \bar{y} = \frac{\Sigma XY}{\Sigma X^2}(x - \bar{x}) \quad \text{where } X = x - \bar{x},\ Y = y - \bar{y}$$

To find ΣX^2, ΣXY the table is given as follows

x	$X = x - \bar{x}$ $= x - 43$	y	$Y = y - \bar{y}$ $= y - 38$	X^2	XY
46	3	40	2	9	6
42	-1	38	0	1	0
44	1	36	-2	1	-2
40	-3	35	-3	9	9
43	0	35	-3	0	0
41	-2	37	-1	4	2
45	2	41	3	4	6
				$\Sigma X^2 = 28$	$\Sigma XY = 21$

$\therefore$ Regression Eqn of y on x is

$$y - 38 = \frac{21^3}{28_4}(x - 43)$$

$$\Rightarrow y = \frac{3x}{4} - \frac{129}{4} + 38$$

$$\Rightarrow y = 0.75x + 5.75$$

For $x = 37 \Rightarrow y = 0.75(37) + 5.75 = 33.5$

$\therefore$ If Q is present he will award 33.5 marks for the 8th performance.

4. The heights are of mothers & daughters are given in the following table. Estimate average height of daughter when the height of mother is 64.5 inches.

Height of mother (x)	62	63	64	64	65	66	68	70
Height of daughter (y)	64	65	61	69	67	68	71	65

Sol. Here we have to find y value at $x = 64.5$, so we need to find Regression eq^n of y on x

Now
$$\bar{x} = \frac{\Sigma x}{n} = \frac{62+63+64+64+65+66+68+70}{8} = 65.25$$

Slly
$$\bar{y} = \frac{\Sigma y}{n} = 67.$$

Since $\bar{x}$ is not a whole number, we use Assume mean method to find Regression eq^n of y on x.

$\therefore$ Regression eq^n of y on x is

$$y - \bar{y} = \frac{r\sigma_y}{\sigma_x}(x - \bar{x})$$

$$\Rightarrow y - \bar{y} = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{n}}{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}(x - \bar{x})$$

$$\Rightarrow y - \bar{y} = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2}(x - \bar{x})$$

x	y	$X = x-65$	$Y = y-67$	X^2	XY
62	64	-3	-3	9	9
63	65	-2	-2	4	4
64	61	-1	-6	1	6
64	69	-1	2	1	-2
65	67	0	0	0	0
66	68	1	1	1	1
68	71	3	4	9	12
70	65	5	-2	25	-10
		$\Sigma X = 2$	$\Sigma Y = -6$	$\Sigma X^2 = 50$	$\Sigma XY = 20$

0.43

∴ Regression eqⁿ of y on x is

$$(y-67) = \frac{8(80)+12}{8(50)-4}(x-65.25)$$

$$y = 67 + \frac{172}{396}(x-65.25)$$

$$y = 0.4343x - 0.4344(65.25) + 67$$

$$\Rightarrow \boxed{y = 0.4343x + 38.66}$$

For $x = 64.5 \Rightarrow y = 0.4343(64.5) + 38.66$

$$\boxed{y = 66.67}$$

∴ Height of daughter when mother height is 64.5 inches = 66.67 inches.

5. Determine the eqⁿ of a straight line which fits the data

X	10	12	13	16	17	20	25
Y	10	22	24	27	29	33	37

Sol. Let the Regression line of Y on X be $Y = a + bX$

then Normal eqⁿs are $\left.\begin{array}{l} na + b\Sigma X = \Sigma Y \\ a\Sigma X + b\Sigma X^2 = \Sigma XY \end{array}\right\} \rightarrow (1)$

X	Y	X^2	XY
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
$\Sigma X = 113$	$\Sigma Y = 182$	$\Sigma X^2 = 1938$	$\Sigma XY = 3186$

sub above values in (1)

we get $7a + 113b = 182$

$113a + 1938b = 3186$

Solving we get $a = 0.82$, $b = 1.56$

∴ Regression eqⁿ of Y on X is $\boxed{Y = 0.82 + 1.56X}$

6. If 'θ' is the angle b/w 2 regression lines and standard deviation of y is twice the S.D of x & $r = 0.25$ then find θ.

Sol.ⁿ Given $\sigma_y = 2\sigma_x$ & $r = 0.25$ = Correlation coeff of x & y.

we know that

$$\tan\theta = \left(\frac{1-r^2}{|r|}\right)\left(\frac{\sigma_x \sigma_y}{\sigma_x^2+\sigma_y^2}\right)$$

$$= \left(\frac{1-0.25^2}{0.25}\right)\left(\frac{2\cancel{\sigma_x^2}}{5\cancel{\sigma_x^2}}\right)$$

$$= 3.75\left(\frac{2}{5}\right) = 1.5$$

$$\therefore \theta = \tan^{-1}(1.5) = 56.30°$$

$$\boxed{\therefore \theta = 56.30°}$$

7. If $\sigma_x = \sigma_y = \sigma$, & the angle between regression lines is $\tan^{-1}\left(\frac{4}{3}\right)$ then find r.

Sol.ⁿ

$$\tan\theta = \left(\frac{1-r^2}{|r|}\right)\left(\frac{\sigma_x \sigma_y}{\sigma_x^2+\sigma_y^2}\right)$$

$$\tan\theta = \left(\frac{1-r^2}{r}\right)\left(\frac{\cancel{\sigma^2}}{2\cancel{\sigma^2}}\right)$$

$$\Rightarrow \tan\theta = \frac{1-r^2}{2r}$$

$$\Rightarrow \theta = \tan^{-1}\left(\frac{1-r^2}{2r}\right) = \tan^{-1}\left(\frac{4}{3}\right) \text{ (given)}$$

$$\Rightarrow \frac{1-r^2}{2r} = \frac{4}{3} \Rightarrow 3 - 3r^2 = 8r$$

$$\Rightarrow 3r^2 + 8r - 3 = 0$$

$$\Rightarrow r = \frac{-8 \pm \sqrt{64+36}}{6} = \frac{-8 \pm 10}{6} = \frac{-8-10}{6} \text{ or } \frac{2}{6}$$

$$= -3 \text{ or } \frac{1}{3}$$

$\boxed{\therefore r = 0.33}$ since $-1 < r < 1$ $\therefore r \neq -3$

8. If $x = 2y+3$ & $y = kx+6$ are the regression lines of x on y and y on x respectively then show that

a) $0 \le k \le \frac{1}{2}$

b) If $k = \frac{1}{8}$ then find $r, \bar{x}, \bar{y}$.

Sol^n Given $x = 2y+3$,

$$\Rightarrow r\frac{\sigma_x}{\sigma_y} = 2.$$

$$\Rightarrow r = \frac{2\sigma_y}{\sigma_x}$$

$y = kx+6$

$$r\frac{\sigma_y}{\sigma_x} = k$$

$$r = \frac{k\sigma_x}{\sigma_y}$$

$$\Rightarrow r^2 = 2k \qquad (\because -1 \le r \le 1)$$

$$\Rightarrow 0 \le 2k \le 1 \qquad \Rightarrow 0 \le r^2 \le 1$$

$$\Rightarrow \boxed{0 \le k \le \frac{1}{2}}$$

If $k = \frac{1}{8} \Rightarrow r^2 = \cancel{2}\left(\frac{1}{\cancel{8}\,4}\right) \Rightarrow r = \pm\frac{1}{2}$

Since $(\bar{x}, \bar{y})$ satisfies the regression eq^ns, we have

$$\bar{x} = 2\bar{y}+3 \rightarrow (1) \quad \& \quad \bar{y} = \frac{1}{8}\bar{x}+6 \rightarrow (2)$$

$$\bar{x} - 2\bar{y} = 3$$

$$\underset{-}{\bar{x}} \underset{+}{-8\bar{y}} = \underset{+}{-48}$$

$$6\bar{y} = 51$$

$$\Rightarrow \bar{y} = \frac{51}{6} \text{ sub in (1), we } \bar{x} = \frac{51}{3}+3 = \frac{60}{3} = 20$$

$$\therefore \bar{x} = 20; \ \bar{y} = \frac{20}{6}$$

9. Calculate the coeff of correlation if the eq^ns of regression lines x on y and y on x are $7x-16y+9=0$ and $5y-4x-3=0$ and mean of x & y

Sol. Given Regression eqn of x on y & y on x are

$7x - 16y + 9 = 0 \Rightarrow x = \frac{16}{7}y + \frac{9}{7} \Rightarrow b_{xy} = \frac{16}{7}$

$-4x + 5y - 3 = 0 \Rightarrow y = \frac{4}{5}x + \frac{3}{5} \Rightarrow b_{yx} = \frac{4}{5}$

solving above eqns we get $(\bar{x}, \bar{y})$ since it satisfies the Regression lines.

~~So solving them we get~~

~~$\bar{x} =$~~ solving them we get $\bar{x} = 0.1034$, $\bar{y} = 0.5172$

~~$\bar{y} =$~~

Correlation coeff $r = \sqrt{b_{xy} \cdot b_{yx}}$

$$= \sqrt{\frac{16}{7} \times \frac{4}{5}} = 0.22$$

10. Given $N = 10$, $\sigma_x = 5.4$, $\sigma_y = 6.2$ & sum of the products of deviations ~~of the~~ from the mean of x & y is 66. then find correlation coeff r.

Soln. we know that $r = \sqrt{b_{xy} \cdot b_{yx}}$

where $b_{yx} = r\frac{\sigma_y}{\sigma_x} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} = \frac{66}{n\sigma_x^2} = \frac{66}{10(5.4)^2}$

$$\therefore b_{yx} = \frac{66}{291.6} = 0.226$$

slly $b_{xy} = r\frac{\sigma_x}{\sigma_y} = \frac{\Sigma XY}{\Sigma Y^2} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2} = \frac{66}{n\sigma_y^2}$

$$\Rightarrow b_{xy} = \frac{66}{10(6.2)^2} = \frac{66}{384.4} = 0.1716.$$

$$\therefore r = \sqrt{(0.226)(0.1716)} = 0.1969$$

$$\boxed{\therefore r = 0.1969}$$

Unit - 3 :

Sampling distributions & Statistical Inferences.

Sampling :

<u>Population</u> : Population is totality of statistical data forming a subject of investigation.

eg: population of heights of Indians etc.

<u>Size of population</u> is no. of observations in population & is denoted by N.

<u>Sampling</u> : The process of selection of a sample is called Sampling.

eg : to assess the quality of a bag of rice, we examine only a portion of it by taking a handful of it ~~from the~~ bag & then decide to purchase it or not.

Thus in estimating the characteristics of the population, instead of enumerating entire population, only the individuals in the sample are examined. Then the sample characteristics are utilised to estimate the population.

To eliminate any possibility of bias in the sampling procedure choose a <u>random sample</u> where observations are made independently & at random.

<u>Sample</u> : It is a subset of population & size of the sample is denoted by 'n'.

<u>Types of Sampling</u>:

1. 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.

   eg: Selecting randomly 20 words from a dictionary.

Note : (a) No. of samples with replacement $= N^n$

(b) " " " without " $= {}^N C_n$.

2. <u>Stratified Random Sampling</u>: The population is first ~~subsubdivided~~ subdivided into several parts (or small groups) called <u>Strata</u> according to some relevant characteristics. Then a sample is selected from each stratum at random.

3. <u>Systematic Sampling</u> (or Quasi-Random Sampling): In this, population is arranged in some order. Then from first 'K' items (say) one unit is selected at random. This unit & every Kth unit combined together form systematic sample. In this method only the first member is chosen at random.

4. <u>Purposive (or Judgement) Sampling</u>: In this method, the units of sample are chosen according to convenience & personal choice of the individual, who selects the sample. This method is suitable when sample is small. Here, the investigator must have a a thorough knowledge of the population. It is always subject to some kind of bias. <u>eg</u>: To select 20 students from a class of 100 to analyse extra-curricular activities of the students, the investigator would select those, who according to him, would represent the class.

5. <u>Sequential Sampling</u>: It consists of a sequence of sample drawn one after another from the population depending on results of previous samples.

ie. If first sample leads to no clear decision, a second sample is drawn, & if required a third sample is drawn to arrive at a final decision to accept or reject the lot.

<u>Classification of Samples</u>:

(1) Large sample: If $n \geq 30$ sample is said to be large sample.

(2) Small sample: If $n < 30$ sample is small sample or exact sample.

<u>Parameter</u> : It is a statistical measure based on all units of a population . eg: Mean μ, Variance σ^2 etc .

<u>Statistic</u> : It is a statistical measure based on all units of a selected sample . eg : sample mean $\bar{x}$, sample Variance s^2 etc .

Note : Value of a statistic varies from sample to sample ($\because$ units of samples are not same), but parameter always remains constant .

SAMPLING DISTRIBUTION of a STATISTIC

It helps us

(1) To estimate unknown population parameter from known statistic .

(2) To set confidence limits of parameter within which the parameter values are expected to lie .

(3) To test a hypothesis & to draw a statistical inference from it .

Central limit theorem

If $\bar{x}$, be mean of a sample size n, drawn from a population with mean μ & S.D σ then sampling dist of sample mean $\bar{x}$ is approx^n normal dist with mean μ, S.D = S.E of $\bar{x}$ provided ($n \geq 30$)

$$Z = \frac{\bar{x}-\mu}{(\sigma/\sqrt{n})}$$ is a random variable whose distribution function approaches that of standard normal distribution $N(z;0,1)$ as $n \to \infty$.

<u>Standard Error</u> (S.E) : It enables us to determine the confidence limits within which the parameters are expected to lie .

<u>Imp. Formulae</u> :

(1) S.E of sample mean $= \dfrac{\sigma}{\sqrt{n}}$

(2) " " " proportion $= \sqrt{\dfrac{PQ}{n}}$, where $Q = 1-P$

(3) " " " Standard deviation $= \dfrac{\sigma}{\sqrt{2n}}$

(4) (a) S.E of difference of 2 sample means $\bar{x}_1$ & $\bar{x}_2$ is $= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

when populations are different.

(b) when samples are from same population then

S.E of difference of 2 sample means $\bar{x}_1$ & $\bar{x}_2$ $= \sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$

(c) when samples are having standard deviations s_1, s_2 for the 2 samples whose means are $\bar{x}_1, \bar{x}_2$ then

S.E of 2 sample means $\bar{x}_1$ & $\bar{x}_2$ $= \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$

Or $= \sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ where

$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2}$.

(5) S.E of difference of 2 sample propotions p_1, p_2

$= \sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}}$ where P_1, P_2 are population propotions

$= \sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ when samples are from same population

$= \sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}$ when population propotions are not given.

$= \sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ where $P = \frac{n_1p_1+n_2p_2}{n_1+n_2}$ &

$Q = 1-P$.

$$\boxed{\text{Correction factor} = \frac{N-n}{N-1}}$$

For finite population of size N, when a sample is drawn without replacement,

(i) S.E of sample mean $\bar{x} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

(ii) " " " propotion $= \sqrt{\frac{PQ}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$.

Sampling Distribution of Mean (σ Known)

(i) Mean of the population $\mu = \frac{\Sigma x_i}{N}$

(ii) Variance " " " $\sigma^2 = \frac{1}{N} \Sigma(x_i - \bar{x})^2$

Do problems.

1pb A population consists of 5, 10, 14, 18, 13, 24. Consider all possible samples of size two which can be drawn without replacement from the population. Find

(a) mean of the population
(b) S.D " " "
(c) mean of sampling distribution of means.
(d) S.D " " " " .

Sol: To make samples of size 2 without replacement :

(5, 10), (5, 14), (5, 18), (5, 13), (5, 24)
(10, 14), (10, 18), (10, 13), (10, 24),
(14, 18), (14, 13), (14, 24),
(18, 13), (18, 24)
(13, 24)

No. of samples $= {}^6C_2 \; (\because {}^NC_n) = \frac{\cancel{6}^{3} \times 5}{\cancel{2}} = 15$.

(a) Mean of population $\mu = \frac{5+10+14+18+13+24}{6} \quad \{\because N=6\}$

$= \frac{84}{6} = 14 \quad \boxed{\therefore \mu = 14}$

(b) S.D of population $\sigma = \sqrt{\sigma^2}$ where

$\sigma^2 = \frac{\Sigma(x_i - \bar{x})^2}{N} = \cancel{\sum_{i=1}^{6}(x_i-\mu)^2}$

$= \frac{(5-14)^2 + (10-14)^2 + (14-14)^2 + (18-14)^2 + (13-14)^2 + (24-14)^2}{5}$

$= \frac{81+16+0+16+1+100}{6}$

$\Rightarrow \sigma^2 = \frac{214}{6} = 35.67$.

$\therefore \sigma = \sqrt{35.67}$

Mean of the samples :

$\frac{5+10}{2}$ ~~7.5~~ $\quad \frac{5+14}{2}$ = ~~9.5~~ $\quad \frac{5+18}{2}$ = ~~11.5~~ $\quad \frac{5+13}{2} \quad \frac{5+24}{2}$

$\frac{10+14}{2} \quad \frac{10+18}{2} \quad \frac{10+13}{2} \quad \frac{10+24}{2}$

$\frac{14+18}{2} \quad \frac{14+13}{2} \quad \frac{14+24}{2}$

$\frac{18+13}{2} \quad \frac{18+24}{2}$

$\frac{13+24}{2}$

ie 7.5 9.5 11.5 9 14.5

12 14 11.5 17

16 13.5 19

15.5 21

18.5

(c) Mean of sampling distribution of mean is $\mu_{\bar{x}}$

$$\mu_{\bar{x}} = \frac{(7.5+9.5+11.5+9+14.5+12+14+11.5+17+16+13.5+19+15.5+21+18.5)}{15}$$

$= 14.$

(d) Variance of S.D of mean is

$$\sigma_{\bar{x}}^2 = \frac{(7.5-14)^2+(9.5-14)^2+(11.5-14)^2+(9-14)^2+(14.5-14)^2+(12-14)^2+(14-14)^2+(11.5-14)^2+(17-14)^2+(16-14)^2+(13.5-14)^2+(19-14)^2+(21-14)^2+(18.5-14)^2}{15}$$

$= \frac{214}{15} = 14.2666$

$\therefore$ S.D of sampling distribution of means is

$\sigma_{\bar{x}} = \sqrt{14.2666} = 3.78.$

(3)(b) Samples of sizes 2 are taken from population 3, 6, 9, 15, 27 with replacement. Find (a) μ (b) σ (c) $\mu_{\bar{x}}$ (d) $\sigma_{\bar{x}}$.

(a) $\mu = \dfrac{3+6+9+15+27}{5} = 12$. $(\because N=5)$

(b) $\sigma^2 = \dfrac{(3-12)^2+(6-12)^2+(9-12)^2+(15-12)^2+(27-12)^2}{5} = \dfrac{360}{5} = 72$

$\therefore \sigma = \sqrt{72}$

(c) samples of size 2 with replacement $= N^n = 5^2 = 25$. They are:

$$\left.\begin{matrix} (3,3) & (3,6) & (3,9) & (3,15) & (3,27) \\ (6,6) & (6,3) & (6,9) & (6,15) & (6,27) \\ (9,9) & (9,3) & (9,6) & (9,15) & (9,27) \\ (15,15) & (15,3) & (15,6) & (15,9) & (15,27) \\ (27,27) & (27,3) & (27,6) & (27,9) & (27,15) \end{matrix}\right] = 25.$$

mean $\mu_{\bar{x}} = ?$

Means of samples are

$$\bar{x}\left\{\begin{matrix} 3 & 4.5 & 6 & 9 & 15 \\ 6 & 4.5 & 7.5 & 10.5 & 16.5 \\ 9 & 6 & 7.5 & 12 & 18 \\ 15 & 9 & 10.5 & 12 & 21 \\ 27 & 15 & 16.5 & 18 & 21 \end{matrix}\right.$$

$$\therefore \mu_{\bar{x}} = \frac{\begin{bmatrix} 3+4.5+6+9+15+6+4.5+7.5+10.5+16.5+9+6+7.5+ \\ 12+18+15+9+10.5+12+21+27+15+16.5+18+21 \end{bmatrix}}{25}$$

$= \dfrac{300}{25} = 12$.

(d)
$$\sigma_{\bar{x}}^2 = \frac{\begin{matrix}(3-12)^2+(4.5-12)^2+(6-12)^2+(9-12)^2+(15-12)^2+(6-12)^2+(4.5-12)^2 \\ +(7.5-12)^2+(10.5-12)^2+(16.5-12)^2+(9-12)^2+(6-12)^2+(7.5-12)^2 \\ +(12-12)^2+(18-12)^2+(15-12)^2+(9-12)^2+(10.5-12)^2+(12-12)^2 \\ +(21-12)^2+(27-12)^2+(15-12)^2+(16.5-12)^2+(18-12)^2+(21-12)^2\end{matrix}}{25}$$

$= \dfrac{909}{25}$

$\sigma_{\bar{x}} = \sqrt{\dfrac{909}{25}} = \underline{6.03}$.

③ A random sample of size 100 is taken from an infinite population having $\mu = 76$ & variance $\sigma^2 = 256$. What is the probability that $\bar{x}$ will be b/w 75 & 78.

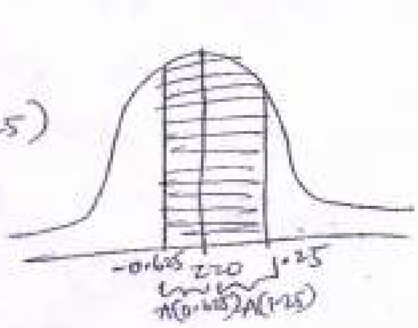
Sol: $n = 100$, $\mu = 76$, $\sigma^2 = 256 \Rightarrow \sigma = 16$.

To find $P(75 \leq \bar{x} \leq 78) = ?$

we have $z = \dfrac{\bar{x} - \mu}{(\sigma/\sqrt{n})}$

$\bar{x}_1 = 75 \Rightarrow z_1 = \dfrac{75 - 76}{16/\sqrt{100}} = -0.625$

$\bar{x}_2 = 78 \Rightarrow z_2 = \dfrac{78 - 76}{16/\sqrt{100}} = 1.25$



$\therefore P(75 \leq \bar{x} \leq 78) = P(-0.625 \leq z \leq 1.25)$

$= A(0.625) + A(1.25)$

$= 0.2324 + 0.3944$

$= 0.628$ (ans).

④ A normal population has mean of 0.1 & S.D of 2.1. Find the probability that mean of a sample of size 900 will be negative.

Sol: $\mu = 0.1$, $\sigma = 2.1$, $n = 900$. To find $P(\bar{x} < 0) = ?$

we have $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{\bar{x} - 0.1}{2.1/\sqrt{900}} = \dfrac{\bar{x} - 0.1}{0.07}$

$\bar{x} - 0.1 = 0.07z \Rightarrow \bar{x} = 0.1 + 0.07z$

$\therefore P(\bar{x} < 0) = P(0.1 + 0.07z < 0) = P(0.07z < -0.1)$

$= P\left(z < -\dfrac{0.1}{0.07}\right)$

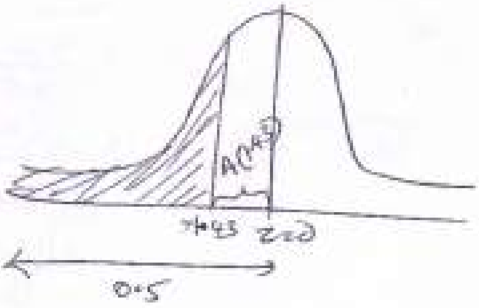
$= P(z < -1.43)$

$= 0.5 - A(1.43)$

$= 0.5 - 0.4236$

$= 0.0764$.

Ans.



ESTIMATION

They are of 2 types: (1) Point Estimation (2) Interval Estimation.

(1) Def^n: If an estimate of the population parameter is given by a single value, then the estimate is called a <u>Point Estimation</u> of the parameter. eg: If height of a student is measured as 162 cms then it gives a point estimation.

(2) If an estimate of the population is given by two different values b/w which the parameter may be considered to lie, then the estimate is called an <u>Interval estimation</u> of the parameter.

eg: If height of student lies b/w 159.5 cm & 166.5 cm then it gives interval estimation.

Confidence interval estimates of parameters

The confidence interval has probability of correctly estimating the true value of the population parameter.

$$\left.\begin{array}{l} \bar{x} \pm 1.96\,(S.E \text{ of } \bar{x}) \\ p \pm 1.96\,(S.E \text{ of } p) \\ (\bar{x}_1-\bar{x}_2) \pm 1.96\,(S.E \text{ of } (\bar{x}_1-\bar{x}_2)) \\ (p_1-p_2) \pm 1.96\,(S.E \text{ of } (p_1-p_2)) \end{array}\right]$$

are 95% confidence limits for
single mean
" proportion
Difference of mean
" " propotion.

Maximum Error of Estimate E:

$$\boxed{E = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)}$$ for large sample

$$\boxed{E = t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)}$$ " small "

Do problems:

Testing of Hypothesis

There are many problems, in which, rather than estimating the value of a parameter we need to decide whether to accept or reject a statement about the parameter.

This statement is called a hypothesis & the decision-making procedure about the hypothesis is called Testing of hypothesis.

ie The procedure which enables us to decide on the basis of sample results whether a hypothesis is true or not, is called Test of Hypothesis or Test of Significance.

Test of Hypothesis involves following steps:

Step 1: Statement (or assumption) of hypothesis;

H_0 (Null Hypothesis) &

H_1 (Alternative Hypothesis).

Step 2: Specification of Level of Significance (L.O.S) 'α':

L.O.S is the max. possibility with which we are willing to risk an error in rejecting H_0 when it is true.

Step 3: Identification of Test Statistic:

say z, t, F etc.

Step 4: Critical region: ie z_α or t_α etc.

Step 5: Making Decision.

If computed value < critical value, we Accept H_0, otherwise reject H_0.

There are 2 types of Hypothesis

(1) Null Hypothesis: Denoted by H_0. It is a definite statement which asserts that there is no significant difference b/w the statistic & the population parameter

eg: $H_0: \mu = \mu_0$

(2) Alternative Hypothesis: It always contradicts H_0 & is denoted by H_1. If $H_0: \mu = \mu_0$ then H_1 would be

(i) $H_1: \mu \neq \mu_0$ (or) (ii) $H_1: \mu > \mu_0$ (or) (iii) $H_1: \mu < \mu_0$

(i) is known as 2-Tailed alternative & (ii) is right tailed

& (iii) is " " left- "

Errors of Sampling : They are

(i) Type I error : Reject H_0 when it is true
(or)
α error
(or)
Producer's risk

(ii) Type II error (or) β error (or) Consumer's risk :
Accept H_0 when it is wrong

Critical Values (Z_α) of Z (ie for Large Samples where $z = \frac{t - E(t)}{S.E(t)}$)

	1% (0.01)	5% (0.05)	10% (0.1)
Two-tailed Test	$\lvert Z_\alpha \rvert = 2.58$	$\lvert Z_\alpha \rvert = 1.96$	$\lvert Z_\alpha \rvert = 1.645$
Right- " "	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
Left- " "	$= -2.33$	$= -1.645$	$= -1.28$

Test of Significance for large samples:

Tests of significance used in large samples are different from those used in small samples, ~~because~~ If S.D of population is not known, it can be replaced by S.D of sample, whereas in small samples it is not possible.

Under Large samples we see 4 important tests of Significance :

(1) Testing of significance for single mean.
(2) " " " " difference of means.
(3) " " " " single proportion.
(4) " " " " difference of proportions.

Note: The confidence limits (or fiducial limits) for single mean are

(i) $\bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}}$ or $\left(\bar{x} - 2.58 \frac{\sigma}{\sqrt{n}}, \bar{x} + 2.58 \frac{\sigma}{\sqrt{n}} \right)$ for 99% confidence

(ii) $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$ for 95% confidence limits ; etc.

Formulae for Testing of Hypothesis -

Test of Hypothesis for single mean :

(i) $z = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ where $\bar{x} \rightarrow$ sample mean

$\mu \rightarrow$ population "

σ — " S.D

n — sample size.

(Here $\left.\begin{matrix}S.E \\ \text{of } \bar{x}\end{matrix}\right\} = \frac{\sigma}{\sqrt{n}}$)

(For Testing of Hypothesis for difference of means :

(i) $z = \dfrac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ where $\bar{x}_1, \bar{x}_2$ — sample means

σ_1, σ_2 — S.D of populations

$n_1, n_2 \rightarrow$ Sample sizes

(Here $\left.\begin{matrix}S.E \\ \text{of } \bar{x}_1-\bar{x}_2\end{matrix}\right\} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$)

(ii) If samples are from same population then

$\sigma_1^2 = \sigma_2^2 = \sigma^2$ &

$z = \dfrac{\bar{x}_1-\bar{x}_2}{\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$. (Here $\left.\begin{matrix}S.E \\ \text{of } \bar{x}_1-\bar{x}_2\end{matrix}\right\} = \sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$)

(ii) If S.D of populations are unknown they can be replaced by S.D S_1, S_2 of samples.

$\therefore \quad z = \dfrac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$ (Here $\left.\begin{matrix}S.E \\ \text{of } \bar{x}_1-\bar{x}_2\end{matrix}\right\} = \sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$)

Note: The confidence limits are

(a) $(\bar{x}_1-\bar{x}_2) \pm 2.58$ (S.E of $\bar{x}_1-\bar{x}_2$) for 99%.

(b) $(\bar{x}_1-\bar{x}_2) \pm 1.96$ (S.E of $(\bar{x}_1-\bar{x}_2)$ for 95%.

For Single Proportion :

$z = \dfrac{p-P}{\sqrt{\frac{PQ}{n}}}$ where $p = \frac{x}{n} \rightarrow$ sample proportion

$P \rightarrow$ Population "

$Q = 1-P$

$n \rightarrow$ sample size.

$x \rightarrow$ no. of successes.

Note: Confidence interval for proportion P are :

$\left(P \pm 3\sqrt{\frac{PQ}{n}}\right)$.

If P is not known we use 'p' (sample proportion)

& limits are $\left(p \pm 3\sqrt{\frac{pq}{n}}\right)$.

For difference of propotions:

(i) $z = \dfrac{p_1 - p_2}{\underbrace{\sqrt{\dfrac{P_1Q_1}{n_1} + \dfrac{P_2Q_2}{n_2}}}_{S.E}}$ where $p_1, p_2 \rightarrow$ sample proportions

$P_1, P_2 \rightarrow$ Population "

$Q_1 = 1 - P_1$, $Q_2 = 1 - P_2$

$n_1, n_2 \rightarrow$ sample sizes.

Also $p_1 = \dfrac{x_1}{n_1}$, $p_2 = \dfrac{x_2}{n_2}$.

Note: When population proportions are unknown use sample proportions. ie, $z = \dfrac{p_1 - p_2}{\underbrace{\sqrt{\dfrac{p_1q_1}{n_1} + \dfrac{p_2q_2}{n_2}}}_{S.E}}$.

(ii) Other method (called method of pooling)

$z = \dfrac{p_1 - p_2}{\underbrace{\sqrt{PQ\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}_{S.E}}$ where $P = \dfrac{n_1P_1 + n_2P_2}{n_1 + n_2} \left(= \dfrac{x_1 + x_2}{n_1 + n_2} \right)$

or $z = \dfrac{p_1 - p_2}{\underbrace{\sqrt{pq\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}_{S.E}}$ where $p = \dfrac{n_1p_1 + n_2p_2}{n_1 + n_2} \left(= \dfrac{x_1 + x_2}{n_1 + n_2} \right)$.

The confidence interval or limits are:

(a) $(p_1 - p_2) \pm z_\alpha \left[S.E \text{ of } p_1 - p_2 \right]$.

Do Problems.

① A truck company claims that average life of certain tyres is atleast 28000 miles. To check the claim it puts 40 of these tyres on its trucks & gets a mean life time of 27463 miles with a S.D of 1348 miles, Can the claim be true?

Sol: $n = 40$, $\bar{x} = 27463$, $s = 1348$, $\mu = 28000$

$= \sigma$ ($\because \sigma$ is not given)

$H_0 : \mu = 28000$ (2 tailed test)

$H_1 : \mu \neq 28000$ (not mentioned)

$\alpha = 0.05$ ($\because$ not mentioned)

$z_\alpha = 1.96$ (for 95% confidence)

$z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{27463 - 28000}{1348/\sqrt{40}} = -2.52$.

$\because |z| = 2.52$ & $|z| > z_\alpha \Rightarrow$ Reject H_0

ie The claim is not true.

Solved Important model questions of Unit-III

1. A population consists of 5, 10, 14, 18, 13, 24. Consider all possible samples of size two which can be drawn without replacement from the population. Find out
(a) Mean of the population (b) Standard deviation of population
(c) Mean of Sampling distribution of means.
(d) S.D of Sampling distribution of means.

Sol. Given population Size $= N = 6$.
Sample Size $n = 2$.
No of samples of size '2' selected from given population (without replacement) $= N_{C_n} = 6C_2 = \frac{6!}{4!\,2!} = \frac{\cancel{6}^{3}\times 5}{\cancel{2}} = 15$

and The samples are. (1)

$\{(5,10),(5,14),(5,18),(5,13),(5,24)$
$(10,14),(10,18),(10,13),(10,24)$
$(14,18),(14,13),(14,24)$
$(18,13),(18,24),(13,24)\}$

(a) Mean of the population $= \mu = \frac{5+10+14+18+13+24}{6} = \frac{84}{6} = 14$

$\boxed{\therefore \mu = 14}$

(b) Standard Deviation of the population $= \sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{n}}$

$\Rightarrow \sigma^2 = \frac{\Sigma(x-\mu)^2}{N} = \frac{(5-14)^2+(10-14)^2+(18-14)^2+(13-14)^2+(24-14)^2}{6}$

$\Rightarrow \sigma^2 = \frac{81+16+16+1+100}{6} = \frac{133+81}{6} = \frac{214}{6} = 35.67$

$\therefore \sigma = \sqrt{35.67} = 5.972$

$\boxed{\therefore \sigma = 5.972}$

c) The means of samples of size 2 are

$\{7.5, 9.5, 11.5, 9, 14.5, 12, 14, 11.5, 17, 16, 13.5, 19, 15.5, 21, 18.5\}$

$\therefore$ the mean of sample means $= \frac{210}{15} = 14$

$\therefore$ The Mean of the Sampling distribution of means $= 14 = \mu_{\bar{x}}$

$\therefore$ we observed that $\boxed{\mu = \mu_{\bar{x}}}$

d) The Variance of sampling distribution of means is

$$\sigma_{\bar{x}}^2 = \frac{(7.5-14)^2 + (9.5-14)^2 + \cdots + (18.5-14)^2}{15}$$

$$= \frac{42.25 + 20.25 + 6.25 + \cdots + 20.25}{15} = 14.2666$$

(2)

$\therefore \sigma_{\bar{x}} = \sqrt{14.266} = 3.78$

$\therefore$ S.D of sampling distribution of means $= 3.78$

Technique

S.NO	sample	Sample mean $\bar{x}$	Mean of Sampling Distribution $\bar{x}-\bar{\bar{x}} = x-14$	$(x-\bar{x})^2$
1	(5,10)	7.5	−6.5	42.25
2	(5,14)	9.5	−4.5	20.25
3	(5,18)	11.5	−2.5	6.25
4	(5,13)	9	−5	25
5	(5,24)	14.5	0.5	0.25
6	(10,14)	12	−2	4
7	(10,18)	14	0	0
8	(10,13)	11.5	−2.5	6.25
9	(10,24)	17	3	9
10	(14,18)	16	2	4
11	(14,13)	13.5	−0.5	0.25
12	(14,24)	19	5	25
13	(18,13)	15.5	1.5	2.25
14	(18,24)	21	7	49
15	(13,24)	18.5	4.5	20.25
Total		$\Sigma x = 210$		214

$\Sigma(x-\bar{x})^2$

2. 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) S.D of a population
c) Mean of the sampling distribution of means.
d) standard Error of Means. (S.D of the sampling distribution of means)

Sol. Given $N = 6$; (population size)
a) Mean of the population $= \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$
b) Variance of the population (σ^2) is

$$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N} = \frac{1}{5}\left[(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2\right]$$

$$\therefore \sigma^2 = 10.8 \Rightarrow \sigma = \sqrt{10.8} = 3.29$$ 3

$\therefore$ S.D of the population $= 3.29$.

c) Number of samples of size '2' from given population

$N^n = 5^2 = 25$

and the samples are given in the table

Sample	Mean of sample (x)	$(x-\bar{x})^2$
(2,2)	2	4
(2,3)	2.5	6.25
(2,6)	4	16
(2,8)	5	25
(2,11)	6.5	0.25
(3,2)	2.5	6.25
(3,3)	3	9
(3,6)	4.5	2.25
(3,8)	5.5	0.25
(3,11)	7	1
(6,2)	4	4
(6,3)	4.5	2.25
(6,6)	6	0
(6,8)	7	1
(6,11)	8.5	6.25

Sample	Mean of Sample	$(x-\bar{x})^2$ $(x-6)^2$
(8,2)	5	1
(8,3)	5.5	0.25
(8,6)	7	1
(8,8)	8	4
(8,11)	9.5	12.25
(11,2)	6.5	0.25
(11,3)	7	1
(11,6)	8.5	6.25
(11,8)	9.5	12.25
(11,11)	11	25
Total	150	135

From the table, we have (4)

$\Sigma x = 150$

$\Rightarrow \mu_{\bar{x}} = \frac{\Sigma x}{n} = \frac{150}{25} = 6$

and $\Sigma(x-\bar{x})^2 = 135$

$\Rightarrow \sigma_{\bar{x}}^2 = \frac{1}{n}\Sigma(x-\bar{x})^2 = \frac{\cancel{135}^{27}}{\cancel{25}_5} = 5.4$

$\therefore \sigma_{\bar{x}} = \sqrt{5.4} = 2.32.$

3. If the population is 3, 6, 9, 15, 27
 a) List all possible samples of size 3 that can be taken without replacement from the finite population.
 b) Mean of Sampling distribution of means
 c) Find S.D of Sampling distribution of means

Sol. No of samples of size 3 ~~from~~ given population $= N_{C_n} = {}^5C_3$

$= \frac{5!}{2!3!} = \frac{5\times4}{2} = 10$ (without Replacement)

The samples are $\{$ (3,6,9), (3,6,15), (3,6,27), (3,9,15), (3,9,27)
(3,15,27), (6,9,15), (6,9,27), (6,15,27)
(9,15,27) $\}$

Sample	Mean of sample	$(x-\bar{x})^2$ $(x-12)^2$
(3,6,9)	6	36
(3,6,15)	8	16
(3,6,27)	12	0
(3,9,15)	9	9
(3,9,27)	13	1
(3,15,27)	15	9
(6,9,15)	10	4
(6,9,27)	14	4
(6,15,27)	16	16
(9,15,27)	17	25
	$\Sigma x = 120$	$\Sigma(x-\bar{x})^2 = 120$

$\therefore$ Mean of sampling distribution of means $= \mu_{\bar{x}} = \frac{\Sigma x}{n} = \frac{120}{10} = 12$

$\therefore$ Variance of sampling distribution of means is

$$\sigma_{\bar{x}}^2 = \frac{1}{n}\Sigma(x-\bar{x})^2 = \frac{1}{10}(120) = 12$$

$$\therefore \sigma_{\bar{x}} = \sqrt{12} = 3.464.$$

<u>CHECK</u> Mean of population $= \frac{\Sigma x}{N} = \frac{3+6+9+15+27}{5} = 12$

Variance of population $= \frac{1}{N}\Sigma(x-\mu)^2$

$$= \frac{1}{5}\left[(3-12)^2+(6-12)^2+(9-12)^2+(27-12)^2\right]$$

$$= \frac{1}{5}\left[81 + 9 + 36 + 225\right] = 70.2 = \sigma^2$$

$$\therefore \sigma = \sqrt{70.2} = 8.3785$$

<u>Relation</u> For finite population, $\mu_{\bar{x}} = \mu = 12$

Variance $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \left(\frac{5-3}{4}\right)\frac{70.2}{3} = \frac{70.2}{6}$

$$\sigma_{\bar{x}} = \sqrt{11.7} = \underline{3.42}$$

4. A normal population has a mean of 0.1 & S.D of 2.1. Find the probability that mean of the sample of size 900 will be negative.

Sol^n sample size = $n = 900$

μ = mean of population = 0.1

σ = S.D of population = 2.1

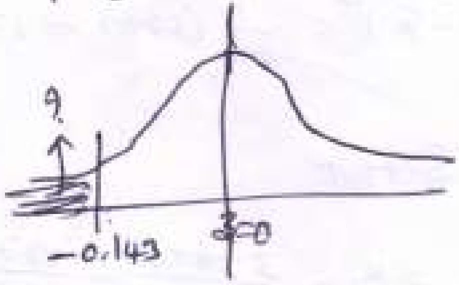
By using Central limit theorem, we have

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

we have to find $P(\bar{x} < 0)$.

For $\bar{x} = 0 \Rightarrow z = \frac{-0.1}{\frac{2.1}{\sqrt{900}}} = -0.1\frac{(30)}{2.1} = -1.43$

$\therefore P(\bar{x} < 0) = P(z < -0.143)$



$\therefore P(\bar{x} < 0) = P(z < -1.43)$

$= 0.5 - P(-1.43 < z < 0)$

$= 0.5 - P(0 < z < 1.43)$

$= 0.5 - 0.4236$

$= 0.0764$

5. A random sample of size 100 is taken from an infinite population having mean $\mu = 76$ & Variance $\sigma^2 = 256$. What is the probability that sample mean will be b/w 75 & 78.

Sol: Given $n = 100$; $\mu = 76$; $\sigma^2 = 256$

$$\Rightarrow \sigma = \sqrt{256} = 16$$

we have to find $P(75 < \bar{x} < 78)$

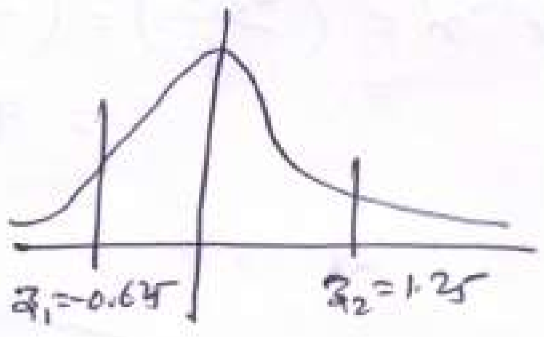
By using central limit theorem,

$$\text{For } \bar{x} = 75,\ z_1 = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{75-76}{16/\sqrt{100}} = -1\left(\frac{10}{16}\right) = -0.625$$

$$\text{For } \bar{x} = 78,\ z_2 = \frac{78-76}{16/10} = 2\left(\frac{10}{16}\right) = 1.25$$

$$\therefore P(75 < \bar{x} < 78) = P(-0.625 < z < 1.25)$$

From normal ~~table~~, curve, we have



$$\begin{aligned} P(75 < \bar{x} < 78) &= P(-0.625 < z < 0) + P(0 < z < 1.25) \\ &= P(0 < z < 0.625) + P(0 < z < 1.25) \\ &= 0.2324 + 0.3944 \\ &= 0.6268 \\ &\simeq 0.63 \end{aligned}$$

6. A sample of size 64 & mean 60 was taken from a population whose S.D is 10. Construct 95% confidence interval for the mean.

Sol.^{n} Given $n=64$; $\bar{x}=60$; $\sigma=10$.

Confidence interval for the mean

$$= \left(\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}},\ \bar{x} + z_\alpha\left(\frac{\sigma}{\sqrt{n}}\right)\right)$$

$$= \left(60 - \frac{10}{\sqrt{64}}(1.96),\ 60 + 1.966\left(\frac{10}{8}\right)\right)$$

$$= (57.55,\ 62.45)$$

7. If Maximum error is 5 & S.D of the population is 80 with 95% confidence, find the sample size

Sol.^{n} Given $E=5$; $\sigma=80$ $z_\alpha=1.96$

$\therefore$ Sample size 'n' $= \left(\frac{z_{\alpha/2}\,\sigma}{E}\right)^2 = \left(\frac{1.96(80)}{5}\right)^2$

$= 983.44$

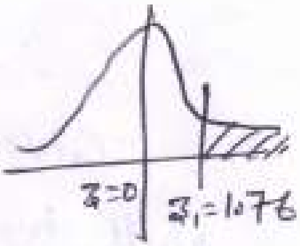
1. A random sample of size 64 is taken from a normal population with $\mu = 51.4$ and $\sigma = 68$. What is the probability that the mean of the sample will be
a) exceed 52.9 b) fall between 50.5 & 52.3 c) be less than 50.6

Sol^n Given $n = 64$
Since sample taken from Normal population,
we use Central limit theorem.
Also $\mu = 51.4$ and $\sigma = 68$

$$\Rightarrow z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}.$$

Now
a) we have to find $P(\bar{x} > 52.9)$

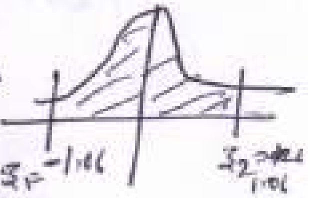
For $\bar{x} = 52.9 \Rightarrow z_1 = \dfrac{52.9 - 51.4}{68/\sqrt{64}} = 1.76$



$$\therefore P(\bar{x} > 52.9) = P(z_1 > 1.76)$$
$$= 0.5 - P(0 < z < 1.76) \text{ (from normal curve)}$$
$$= 0.0392$$

b) we have to find $P(50.5 \le \bar{x} \le 52.3)$

For $\bar{x}_1 = 50.5 \Rightarrow z_1 = \dfrac{50.5 - 51.4}{68/\sqrt{64}} = -1.06$

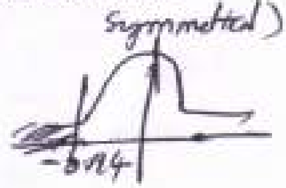


$\bar{x}_2 = 52.3 \Rightarrow z_2 = \dfrac{52.3 - 51.4}{68/\sqrt{64}} = 1.06$

$$\therefore P(50.5 \le \bar{x} \le 52.3) = P(-1.06 < z < 1.06)$$
$$= 2P(0 < z < 1.06) = 0.7108$$
(Since normal curve is symmetrical)

c) we have to find $P(\bar{x} < 50.6)$

For $\bar{x} = 50.6 \Rightarrow z_1 = \dfrac{50.6 - 51.4}{68/\sqrt{64}} = -0.94$



$$P(\bar{x} < 50.6) = P(z < -0.94)$$
$$= 0.5 - P(0 < z < 0.94) = 0.5 - 0.3264 = 0.1736$$

2. A die is tossed 256 times and its turnup with an even digit 150 times. Is the die biased?

Sol: Probability of getting even digit in a single throw of a die (i.e 2 or 4 or 6) $= \frac{3}{6} = \frac{1}{2} = p$

$\therefore q = 1-p = 1-\frac{1}{2} = \frac{1}{2}.$

$\therefore$ Mean $= np = 256\left(\frac{1}{2}\right) = 128 = \mu$

$\sigma^2 =$ Variance $= npq = 256\left(\frac{1}{2}\right)^2 = \frac{256}{4} = 64$

$\Rightarrow$ S.D $= 8 = \sigma$

$\therefore \mu = 128, \quad \sigma = 8$

Given $x =$ no of success $= 150$

Null hypothesis H_0: Die is unbiased

Alternative hypothesis H_1: Die is biased

Level of significance: $\alpha = 0.05$

Test statistic

$$z_{cal} = \frac{x-\mu}{\sigma} = \frac{150-128}{8} = \frac{22}{8} = 2.75$$

Since at 5% level, $z_{tab} = 1.96$

$\therefore z_{cal} > z_{tab} \Rightarrow$ Null hypothesis Rejected

$\Rightarrow$ Alternative hypothesis accepted

$\therefore$ The Die is biased.

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.

Soln Given $n = 400$; $\sigma = 10$; $\bar{x} = 40$

1. Null hypothesis H_0: $\mu = 38$

2. Alternative hypothesis H_1: $\mu \neq 38$

3. Level of significance: $\alpha = 0.05$

4. Test statistic

$$z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{40 - 38}{10/\sqrt{400}} = 4$$

since z_{tab} at $\alpha = 0.05 = 1.96$

$z_{cal} > z_{tab}$, we reject the null hypothesis H_0

i.e the sample is not taken from the population with mean $\mu = 38$. 6

95% Confidence Interval is given by

$$= \left(\bar{x} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right), \bar{x} + 1.96\left(\frac{\sigma}{\sqrt{n}}\right)\right)$$

$$= \left(40 - 1.96\left(\frac{10}{\sqrt{400}}\right), 40 + 1.9\left(\frac{1}{2}\right)\right)$$

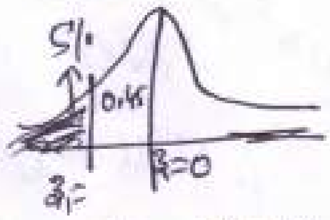
$$= (40 - 0.98, 40 + 0.98)$$

$$= (39.02, 40.98)$$

4. An ambulance service claims that it takes on the average less than 10 mins to reach its destination in emergency calls. A sample of 36 calls has a mean of 11 mins and the variance of 16 mins. Test the claim at 0.05 level of significance.

Sol^n Given $n = 36$ (large sample)

$\bar{x} = 11$ mins ; $s^2 = 16 \Rightarrow s = 4$ mins



Normal curve

1. Null Hypothesis H_0 : $\mu = 10$ mins
2. Alternative Hypothesis H_1 : $\underline{\mu < 10 \text{ mins}}$
3. level of significance : $\alpha = 0.05$
4. Test statistic

$$z_{cal} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{11-10}{4/\sqrt{36}} = \frac{6}{4} = 1.5$$

Since the problem is of one-tail test

From normal curve

$$p(0 < z < z_{tab}) = 0.45$$
$$\Rightarrow z_{tab} = 1.645$$

$\therefore z_{cal} < z_{tab} \Rightarrow$ Accept Null Hypothesis H_0

$\therefore$ An ambulance service takes on average to reach its destination = 10 mins.

5. In 64 randomly selected hours of production, the mean & S.D of the number of acceptance pieces produced by an automatic stamping machine are 1.038 and 0.146. At the 0.05 level of significance does this enable us to reject the Null hypothesis $\mu = 1$ against the alternative hypothesis $\mu > 1$?

Sol. Given $n = 64$, $\bar{x} = 1.038$; $S = 0.146$.

1. Null hypothesis $H_0 : \mu = 1$
2. Alternative hypothesis $H_1 : \mu > 1$ (one-tail test)
3. Level of significance : $\alpha = 0.05$
4. Test statistic :

$$z_{cal} = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{1.038 - 1}{\frac{0.146}{\sqrt{64}}} = 2.082$$

we know that $z_{tab} = 1.645$ (for one-tailed test)

Since $z_{cal} > z_{tab}$, we reject H_0

$\therefore \mu > 1$

i;e Mean of the population is greater that 1.

6. A company claims that its bulbs are superior to those of its main competitor. If a study showed that a sample of 40 of its bulbs have a mean life time of 647 hrs. of continuous use. with a standard deviation of 27 hrs. Test the significance between the difference of two means at 5% level if a sample of 40 bulbs made by main competitor had a mean life of 638 hrs of continuous use with S D of 31 hrs.

Sol Given $n_1 = 40$; $\bar{x}_1 = 647$, $S_1 = 27$ hrs
$n_2 = 40$; $\bar{x}_2 = 638$, $S_2 = 31$ hrs

1. Null Hypothesis $H_0 : \mu_1 = \mu_2$
2. Alternative Hypothesis $H_1 : \mu_1 > \mu_2$

3. Level of Significance: $\alpha = 0.05$

4. Test Statistic

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{647-638}{\sqrt{\frac{(27)^2}{40} + \frac{(31)^2}{40}}} = \frac{9}{\sqrt{\frac{729+961}{40}}}$$

$$\therefore z_{cal} = \frac{9}{6.5} = 1.38$$

Since $z_{tab} = 1.645$ ($\because$ one tailed test)

$\because z_{cal} < z_{tab}$, we accept Null hypothesis H_0

i.e $\mu_1 = \mu_2$

Implies that the difference b/w two sample means is not significant.

7. In a certain factory there are two independent processes for manufacturing the same item. The average weight in a sample of 700 items produced from one process is found to be 250gms with S.D of 30gms while in a sample of 300 items from other process are 300 and 40. Is there Significant difference b/w the means at 1% level

Sol. Given $n_1 = 700$; $\bar{x}_1 = 250$gms ; $s_1 = 30$gms

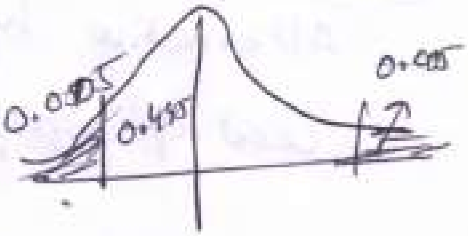
$n_2 = 300$; $\bar{x}_2 = 300$gms; $s_2 = 40$gms

Let the Null hypothesis be $H_0: \mu_1 = \mu_2$

Alternative hypothesis $H_1: \mu_1 \neq \mu_2$ (Two tailed test)

Level of Significance : $\alpha = 0.01$

Test Statistic



$$z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = \frac{250-300}{\sqrt{\frac{30^2}{700}+\frac{40^2}{300}}}$$

$$\Rightarrow z_{cal} = \frac{-50}{\sqrt{\frac{9}{7}+\frac{16}{3}}} = \frac{-50}{\sqrt{\frac{27+112}{21}}} = -50\sqrt{\frac{21}{139}} = -19.823$$

But z_{tab} at 1% level of significance from Normal curve, $P(0<z<z_{tab}) = 0.495$

$\therefore z_{tab} = 2.58$

Since $|z_{cal}| > |z_{tab}|$

we reject the Null hypothesis H_0

$\therefore$ There is a significant difference between the means at 1% level of significance

8. Samples of students were drawn from two universities and from their weights in Kgs, Mean and S.D are calculated and shown below. Test the significance of the difference between the means.

	Mean	S.D	Size
A	55	10	400
B	57	15	100

Sol. Given $n_1 = 400$; $\bar{x}_1 = 55$; $s_1 = 10$
$n_2 = 100$, $\bar{x}_2 = 57$; $s_2 = 15$

Let Null hypothesis be there is no significant difference b/w the means i.e $H_0: \mu_1 = \mu_2$.

Alternative hypothesis $H_1: \mu_1 \neq \mu_2$

Level of significance $\alpha = 0.\cancel{0}1$ (10% level or 95% confidence)

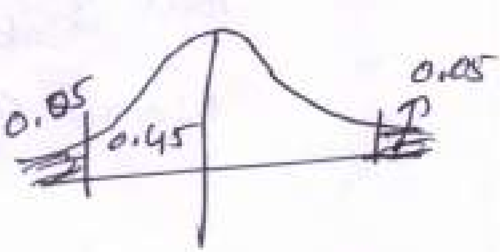
Test statistic

$$z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{55-57}{\sqrt{\frac{100}{400} + \frac{225}{100}}} = \frac{-2}{\sqrt{\frac{1}{4}+\frac{9}{4}}} = \frac{-4}{\sqrt{10}} = -1.26$$

0.45

Since 10% level of significance (two tailed test)

By using Normal curve, we get



$$P(0 < z < z_{tab}) = 0.45$$

$$\Rightarrow z_{tab} = 1.645$$

$\because |z_{cal}| < z_{tab}$, we accept Null Hypothesis H_0

$\Rightarrow$ There is no significant difference b/w the means at 10% level of significance.

9. A manufacturer claimed that atleast 95% of the equipment which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at 5% level of significance. Also find its confidence interval at 5% level. [Confidence interval $= (p \mp 1.645\sqrt{0.0156})$ $= (0.8843, 0.9356)$]

Sol. Given the claim that atleast 95% of the equipment supplied to a factory is conformed to specifications

[Use formula $\left(p - z_{\frac{\alpha}{2}}\left(\sqrt{\frac{pq}{n}}\right), p + z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}\right)$ = Confidence interval at 5% level]

Also $n = 200$

Proportion of faulty items in a sample $= \frac{18}{200}$

$\Rightarrow$ Proportion up to specifications in a sample $= 1 - \frac{18}{200}$

$= \frac{182}{200}$

$\therefore p = \frac{91}{100} = 0.91$

Let Null hypothesis be H_0: $P = 0.95$

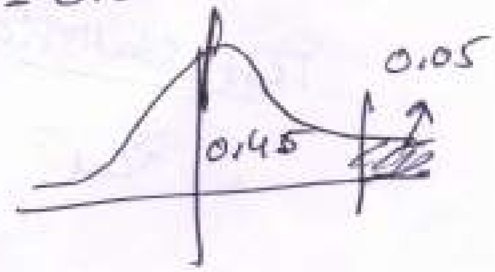
Alternative hypothesis H_1: $P \neq 0.95$ (one-tailed test)

Level of significance: $\alpha = 0.05$

Test Statistic

$$z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

Here $Q = 1 - P = 1 - 0.95 = 0.05$



$$\therefore z_{cal} = \frac{0.91 - 0.95}{\sqrt{\frac{(0.95)(0.05)}{200}}}$$

$$\Rightarrow z_{cal} = \frac{-0.04}{\sqrt{\frac{0.049}{200}}} = \frac{-0.04}{\sqrt{0.000245}} = \frac{-0.04}{0.0156} = -2.564$$

At 5% level of significance (one-tailed)

By Normal curve we have

$z_{tab} = 1.645$

$\because |z_{cal}| > z_{tab}$ we reject H_0

$\Rightarrow \quad P \neq 0.95$

$\Rightarrow P < 0.95 \Rightarrow$ Manufacturer's claim is rejected.

10. In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers?

Sol. Given $n = 600$

Proportion of smokers in a sample $= \dfrac{\overset{\overset{13}{\cancel{65}}}{\cancel{325}}}{\underset{\underset{24}{\cancel{120}}}{\cancel{600}}} = \dfrac{13}{24} = 0.5417$

Let the Null hypothesis be H_0: $P = \frac{1}{2}$

i.e majority of men in a city are not smokers.

Alternative Hypothesis H_1: $P > \frac{1}{2}$ (one tailed)

i.e Majority of men in a city are smokers,

Level of significance: $\alpha = 0.05$

Test statistic

$$z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \quad \text{where } Q = \frac{1}{2}$$

$$= \frac{0.5417 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{600}}} = \frac{30}{0.5}(0.0417)$$

$$= 2.04$$

At 5% level of significance,

$z_{tab} = 1.645$

$\therefore z_{cal} > z_{tab} \Rightarrow$ we reject H_0 and we conclude that the Majority of men in the city are smokers.

11. Experience had shown that 20% of a manufactured product is of the top quality. In one days production of 400 articles only 50 are of top quality. Test the hypothesis at 0.05 level.

Sol. Given Proportion of top qualities in population $= P = \frac{20}{100} = 0.2$.

$n = 400$; Proportion of top quality in sample is

$$p = \frac{50}{400} = 0.125$$

1. let Null hypothesis be $H_0: P = 0.2$. then $Q = 0.8$
2. Alternative Hypothesis $H_1: P \neq 0.2$
3. Level of Significance; $\alpha = 0.05$
4. Test statistic

$$z_{cal} = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.125-0.2}{\sqrt{\frac{(0.2)(0.8)}{400}}} = \frac{-0.075}{0.02} = -3.75$$

Since $z_{tab} = 1.96$ at $\alpha = 0.05$, we have

$\therefore |z_{cal}| > |z_{tab}|$ implies that Reject the Null hypothesis, is rejected

So, we conclude that experience claim is not correct.

12. In two large populations, there are 30% and 25% respectively of fair haired colour. people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations.

Sol. Given Proportion of fair haired in first population $= \frac{30}{100} = 0.3 = P_1$

Proportion of fair haired in second population

$$= \frac{25}{100} = \frac{1}{4} = 0.25 = P_2$$

Given $n_1 = 1200$ & $n_2 = 900$

Let 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 $P_1 = P_2$

Alternative Hypothesis H_1: $P_1 \neq P_2$

Level of significance: $\alpha = 0.05$

Test statistic

$$Z_{cal} = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}} = \frac{0.3 - 0.25}{\sqrt{\frac{0.3(0.7)}{1200} + \frac{0.25(0.75)}{900}}} = 2.56$$

we know Z_{tab} at 5% level $= 1.96$

since $Z_{cal} > Z_{tab}$, we reject H_0 and we conclude that the differene in population proportions is unlikely that the real difference will be hidden.

13. 100 articles from a factory are examined and 10 are found to be defective. 500 similar articles from a second factory are found to be 15 defective. Test the significance b/w the difference of the two proportions at 1% level.

Sol. Given $n_1 = 100$; No of articles defective in first sample $= 10$

$\therefore$ proportion of defective items in first sample $= P_1 = 0.1$

$n_2 = 500$; proportion of defective items in second sample $= \frac{15}{500} = \frac{3}{100} = 0.03 = P_2$

Null Hypothesis $H_0: P_1 = P_2$

Alternative Hypothesis $H_1: P_1 \neq P_2$

Level of significance: $\alpha = 0.01$

Test statistic

$$Z_{cal} = \frac{P_1 - P_2}{\sqrt{qp\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{(Method of pooling)}$$

where $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{100(0.1) + 500(0.03)}{600} = \frac{25}{600} = 0.042$

$q = 1 - p = 1 - 0.042 = 0.958$

$$\therefore Z_{cal} = \frac{0.1 - 0.03}{\sqrt{(0.042)(0.958)\left(\frac{1}{100} + \frac{1}{500}\right)}} = \frac{0.07}{0.022} = 3.18$$

At 1% level, $Z_{tab} = 2.58$

$\therefore Z_{cal} > Z_{tab}$ we reject H_0 $\Rightarrow$ There is significant difference b/w two proportions

14. In an investigation on the machine performance the following results are obtained.

	Inspected	Defects
Machine A	375	17
B	450	22

Test whether there is any significant performance of two machines at $\alpha = 0.05$.

Sol^n Given $n_1 = 375$; $n_2 = 450$; $p_1 = \frac{17}{375} = 0.0453$; $p_2 = \frac{22}{450} = 0.049$

Null Hypothesis H_0: $P_1 = P_2$

Alternative Hypothesis H_1: $P_1 \neq P_2$

Level of significance: $\alpha = 0.05$

Test Statistic

$$z_{cal} = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{(Method of pooling)} \quad \left[\because p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{39}{825} = 0.0472\right]$$

$$= \frac{0.0453 - 0.049}{\sqrt{(0.0472)(0.953)\left(\frac{1}{375} + \frac{1}{450}\right)}}$$

$$= \frac{0.045 - 0.049}{0.015} = -0.267$$

$\therefore |z_{cal}| = 0.267$

Since $|z_{cal}| < z_{tab} = 1.96$, we accept H_0

$\Rightarrow$ There is no significant performance of two machines at 5% level of significance

Unit - 4. Exact Sampling Distributions (Small Samples)

When the sample is small, the Sampling Distribution in many cases may not be normal. Here the test statistics will change.

Degrees of Freedom: denoted by ν. It is the no. of values in a set of data which may be assigned arbitrarily or, It refers to the no. of independent constraints in a set of data.

In general, the no. of degrees of freedom = to the total no. of observations ~~...~~ − no. of independent constraints imposed on the observations.

In brief: The no. of independent variates which make up the statistic is known as degrees of freedom (d.f).

t-distribution (OR) Student's t - Dist.

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} \quad \text{with } \nu = n-1 \text{ d.f.}$$

where $\bar{x}$ - sample mean
μ - Population "
n - sample size

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

Prob. density fn: $f(t) = y_0\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$ where y_0 is a constant got by $\int_{-\infty}^{\infty} f(t)\,dt = 1$.

t-distri. is extensively used in hypothesis about one mean, or about equality of two means when σ is unknown.

Applications of t-Distribution.

(1) To test the significance of the sample mean, when population variance (σ^2) is not given.

(2) To test the significance of the mean of the sample, ie to test if the sample mean differs significantly from the population mean (μ).

(3) To test the significance of the difference between two sample means or to compare 2 samples.

(4) To test the significance of an observed sample correlation coefficient & sample regression coefficient.

Chi-Squared (χ^2) Distribution:

Chi-Squared
It is a conti prob. distri of a conti r.v X with prob. density fn given by

$$f(x) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} \cdot e^{-x/2}, & \text{for } x > 0 \\ 0 & \text{, otherwise} \end{cases}$$

χ^2-distri. was extensively used as a measure of goodness of fit & to test the independence of attributes.

Note ① χ^2 varies from 0 to ∞.

② If χ_1^2 & χ_2^2 are 2-independent distributions with ν_1 & ν_2 dof, then $\chi_1^2 + \chi_2^2$ will be chi-sqrd distri. with $(\nu_1+\nu_2)$ dof. ie It is additive.

★ It is used in sampling distribution analysis of variance, mainly it is used as a measure of goodness of fit & in analysis of r x c tables.

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{\sigma^2} \quad \left\{ \because s^2 = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n-1} \right.$$

— is a value of a r.v having χ^2 distribution with $\nu = n-1$ dof.

Note: (i) Exactly 95% of χ^2 distri. lies betn $\chi^2_{0.975}$ & $\chi^2_{0.025}$

(ii) & when σ^2 is too small χ^2 falls to rt. of $\chi^2_{0.025}$ &

(iii) when σ^2 is too large χ^2 falls to the left of $\chi^2_{0.975}$.

Thus when σ^2 is correct, χ^2 values are to the left of $\chi^2_{0.975}$ or to the right of $\chi^2_{0.025}$.

F-distribution (S.D. of the ratio of 2 sample variances).

It is another imp conti. prob. distri.

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2},$$

where S_1^2, S_2^2 are variances of 2 random samples of sizes n_1 & n_2 with variances σ_1^2 & σ_2^2.

which follows F-distri. with $\nu_1 = n_1-1$ & $\nu_2 = n_2-1$ dof.

Note ① F determines whether the ratio of 2 sample variances S_1 & S_2 is too small or too large.

② when F is close to 1, S_1, S_2 are almost same.

③ F is always a +ve no.

San. Dis of F is of the form $f(F) = K \frac{F^{(\nu_1-2)/2}}{(\nu_1 F+\nu_2)^{(\nu_1+\nu_2)/2}}$

where K is determined by $\int_0^\infty f(F)\,dF = 1.$

t - distribution.

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}} \quad \text{where} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 .$$

p.d.f $\quad f(t) = y_0 \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$

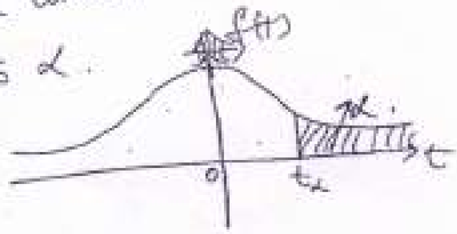
where $\nu = n-1$ dof.

y_0 is a constant got by $\int_{-\infty}^{\infty} f(t)\,dt = 1$.

Note: The mean of std ND as well as t-D is zero, but variance of t-D depends on parameter ν.

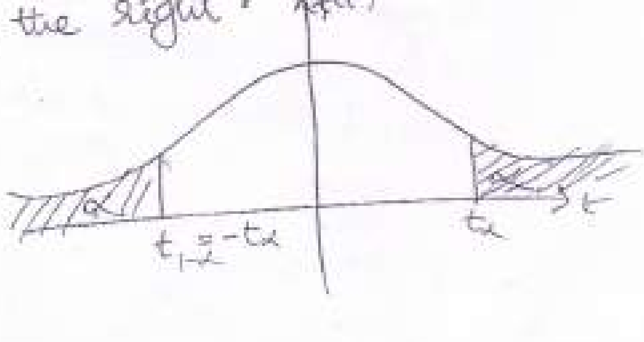
— t-D with ν dof approaches std ND as $\nu \to \infty$.

— t_α denotes area under t-D to its right is equal to α.



★ $t_{1-\alpha} = -t_\alpha$

ie the t-value ~~leaving~~ leaving an area of $1-\alpha$ to the right & ∴ an area α to its left = −ve t-value which leads an area α in the right.



Not needed

t-D is extensively used in hypothesis about one mean, or about equality of 2 means when σ is unknown.

Applications:

1) To test significance of sample mean, when pop. var. is not given.

2) To test significance of the mean of the sample i.e. to test if sample mean differs significantly from pop. mea.

3) To Test significance of difference betn 2 sample means or to compare 2 Samples.

4) -

Chi-Squared (χ^2) Distribution:

p.d.f is given by

$$f(x) = \begin{cases} \dfrac{1}{2^{\nu/2}\,\Gamma(\nu/2)}\, x^{(\nu/2)-1} e^{-x/2}, & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Properties:

(1) χ^2 is not symmetrical, lies entirely in I Quadrant & hence not a normal curve, $\because \chi^2$ varies from 0 to ∞.

(2) It depends only on ν.

— Mean = ν, Variance = 2ν

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{\sigma^2} \quad \left\{ \because s^2 = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n-1} \right.$$

Note: (i) Exactly 95% of χ^2-D lies between $\chi^2_{0.975}$ & $\chi^2_{0.025}$

(ii) When σ^2 is too small, χ^2 falls to the right of $\chi^2_{0.025}$

(iii) When σ^2 is too large, χ^2 falls to the left of $\chi^2_{0.975}$.

When σ^2 is correct χ^2-values are to the left of $\chi^2_{0.975}$ or to the right of $\chi^2_{0.025}$

Applications:

(1) To test goodness of fit.

(2) " " independence of attributes.

(3) " " homogeneity of independent estimation of populalation variance (σ^2)

(4) To test homogeneity of independent estimation of population correlation Coefficient.

F-Distribution

(Sampling Distri. of ratio of 2 Sample Variances)

Not needed. To determine whether the 2 samples come from 2 populations having equal variances.

$$F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} = \frac{\sigma_2^2 s_1^2}{\sigma_1^2 s_2^2}$$

with $\nu_1 = n_1 - 1$ & $\nu_2 = n_2 - 1$ dof

Note: F-D can be used to test equality of several population means, comparing sample variances & analysis of variance.

$F = \frac{S_1^2}{S_2^2}$ under the assumption that 2 normal populations have the same variance. ie $\sigma_1^2 = \sigma_2^2$.

★ F determines whether the ratio of 2 sample variances S_1 & S_2 is too small or too large.

When F is close to 1, the 2 sample variances S_1 & S_2 are almost same.

note: In practice, it is customary, to take the larger sample variance as the numerator

– F is always a +ve no.

The sampling Distribution of F is of the form $f(F) = K \frac{F^{(\nu_1-2)/2}}{(\nu_1 F + \nu_2)^{(\nu_1+\nu_2)/2}}$ where

K is determined by $\int_0^\infty f(F)\,dF = 1$.

Properties.

1) F–D curve lies entirely in first-quadrant.
2) The F-curve depends not only on ν_1, ν_2 but also on the order in which they are used.
3) $F_{1-\alpha}(\nu_1, \nu_2) = \dfrac{1}{F_\alpha(\nu_2, \nu_1)}$ where

$F_\alpha(\nu_1, \nu_2)$ is the value of F with ν_1, ν_2 dof $\ni$ area under F–D curve to the right of F_α is α.

4) mode of F–D is less than unity.

Not Needed

Sampling Distribution of Proportions: (say SDP)

Let 'p' be prob. of occurrence of an event (called success) &
$q = 1 - p$ is prob. of non-occurrence (called failure)

Draw all possible samples of size 'n' from an infinite population.

Compute P proportion of success for each of these samples.

Mean μ_P & variance σ_P^2 of ~~S D of t~~ SDP are

$$\mu_P = p, \quad \sigma_P^2 = \frac{pq}{n} = \frac{p(1-p)}{n}$$

Note: For finite population (without replacement) of size N

$$\mu_P = p, \quad \sigma_P^2 = \frac{pq}{n}\left(\frac{N-n}{N-1}\right).$$

Unit IV

Test of Significance for Small Samples.

On the basis of sample results, the tests of significance enable us to decide, if

(i) the deviation b/w Observed sample statistic & hypothetical parameter value is significant.

(ii) the deviation b/w 2 sample statistics is significant.

Some imp. tests for small samples are:

(a) Students' 't'-test

(b) F-test

(c) χ^2-test

Student's 't' test:

Assumptions of 't'-test:

(i) Sample size, $n < 30$

(ii) The parent population from which sample is drawn is Normal.

(iii) The population standard deviation (may or may not be known) is unknown.

(iv) The sample observations are independent. ie sample is random.

Uses: This test is used

1. to test for a specified mean
2. to test for equality of 2 means (of 2 independent samples drawn from 2 normal populations), S.D of the populations being unknown
3. to test the significance of difference b/w the means of paired data.

let $\bar{x}$ = mean of a sample
n = size of the "
s = S.D of " "
μ = Mean of the population supposed to be normal.

Student's t is defined by the statistic $\boxed{t = \dfrac{\bar{x}-\mu}{SD/\sqrt{n-1}}}$

case(i) When S.D of sample ie 's' is given directly then

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$$

case (ii) When 's' is not given but samples are given then calculate 's' from $\sigma^2 = \dfrac{\sum_i (x_i - \bar{x})^2}{n}$

& then use in the formula above.

note: Also if $S^2 = \dfrac{\sum(x_i-\bar{x})^2}{n-1}$ then $t = \dfrac{\bar{x}-\mu}{S/\sqrt{n}}$

where S^2 is called unbiased estimate of population variance σ^2.

In general while solving problems we ~~deale~~ use σ for s

★ The Confidence or Fiducial limits for μ are

$\boxed{\bar{x} \pm t_{0.05} \cdot \dfrac{S}{\sqrt{n}}}$ at 5% Level of significance

& for $(n-1)$ degrees of freedom.

In general : $\bar{x} \pm t_{\alpha} \dfrac{S}{\sqrt{n}}$

Students 't' Test for Single Mean.

To test the hypothesis that population mean μ has a specified value μ_0 when population S.D σ is <u>not Known.</u>

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

$t = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}}$ where 's' is sample S.D with

$\nu = (n-1)$ degrees of freedom.

$\alpha = 0.05$ (ie 95% confidence limits).

Find table value of 't' at α level of significance.

If calculated value of $t, |t| >$ table value of t

$\Rightarrow$ Reject H_0

If $|t| < t_{\text{table value}} \Rightarrow$ Accept H_0.

Note: For a 2-tailed test, value of $\frac{\alpha}{2}$ is taken for α.

Do problems.

Student's t-test for Difference of Means.

Let μ_1, μ_2 be Means of two Normal Populations

Let n_1, n_2 be the sample sizes from the two populations.

Let $\bar{x}, \bar{y}$ be the means of these 2 independent samples whose S.Ds are s_1, s_2 respectively.

To Test whether the 2 population means are equal.

(ie whether the difference $\mu_1 - \mu_2$ is significant).

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$\nu = (n_1 + n_2 - 2)$ degrees of freedom.

Here $S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$.

Note: If s_1, s_2 are not given then

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2\right] \text{ where}$$

$$\bar{x} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i \quad ; \quad \bar{y} = \frac{1}{n_2}\sum_{i=1}^{n_2} y_i$$

Obtain table value of 't' for 'α' d.f.

If $|t| < t_\alpha \Rightarrow$ Accept H_0

otherwise Reject H_0.

* Confidence limits for difference of 2 population mean are $(\bar{x} - \bar{y}) \pm t_\alpha S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

Note: Same procedure is followed whether samples are taken from different populations or same.

Paired Sample t-Test:

Suppose a business concern is interested to know whether a particular media of promoting sales of a product is really effective or not. Here we have to test whether the average sales "before" and "after" the sales promotion are equal. Th

Thus the particular media is the Experimental unit & the 2 populations are ~~average~~ sales "before" and "after". Hence the two samples are not independent. These are paired observations which arise in many practical situations where each homogeneous experimental unit receives both population conditions.

If $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$ be pairs of Sales data "before" & "after" the sales promotion in a business concern, 'paired t-Test' is applied to test the significance of the difference of the two situations.

Let $d_i = x_i - y_i$ (or $y_i - x_i$) for $i = 1, 2, 3, \ldots n$

$H_0 : \mu_1 = \mu_2$ (ie $\mu = 0$)

$H_1 : \mu_1 \neq \mu_2$

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.

$$t = \frac{\bar{d}}{s/\sqrt{n}} \quad \left\{ \because \mu = 0 \text{ , otherwise } t = \frac{\bar{d} - \mu}{s/\sqrt{n}} \right\}.$$

$$\bar{d} = \frac{1}{n} \sum d_i \quad , \quad s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2$$

α = Level of Significance (L.O.S).

$\nu = n-1$ d.f.

If $|t| < t_\alpha \Rightarrow$ Accept H_0 otherwise Reject H_0

Snedecor's F-Test.

If Variances of the populations are not equal, a significant difference in the means may arise. Hence before we apply t-test for significance of difference of two means, we have to test for equality of population variances using F-test.

the test statistic is $F = \frac{\text{Greater Variance}}{\text{Smaller Variance}}$.

when F is close to 1, the 2 sample variances (say S_1, S_2) are nearly same.

$F_\alpha(\nu_1, \nu_2)$ is value of F with ν_1, ν_2 d.f such that the area under F-distribution to the right of F_α is α.

Clearly value of F at 5% significance is lower than at 1%.

To test the hypothesis that the two population variances σ_1^2 & σ_2^2 are equal:

Let two independent random samples of sizes n_1, n_2 be drawn from 2 normal populations.

The estimates of σ_1^2, σ_2^2 are given by

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{\Sigma(x_i - \bar{x})^2}{n_1 - 1};$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{\Sigma(y_i - \bar{y})^2}{n_2 - 1}$$

where s_1^2, s_2^2 are variances of the two samples.

$$F = \frac{S_1^2}{S_2^2} \quad \left(\text{or } \frac{S_2^2}{S_1^2}\right) \text{ according as } S_1^2 > S_2^2 \ (\text{or } S_2^2 > S_1^2)$$

$\nu_1 = n_1 - 1$; $\nu_2 = n_2 - 1$ are degrees of freedom

If $F_{\text{calculated value}} < F_{\text{tabulated value}}$ we accept H_0 &

conclude σ_1^2, σ_2^2 are equal. Otherwise Reject H_0.

(Do problems)

Chi- Square Test : χ^2-Test .

χ^2 Test is used to test whether differences b/w Observed & expected frequencies are significant . It is mainly used

(1) To Test the goodness of fit

(2) To test the independance of attributes.

(3) To test if the population has a specified value of the variance σ^2 .

In general, the results obtained in an experiment do not agree exactly with the theoretical result s. The magnitude of discrepancy b/w the theory & Observation is given by the quantity χ^2 (a Greek letter, read as chi- square) ,

If $\chi = 0$, observed & expected frequencies coincide completely.

As χ^2 value increases, the discrepancy b/w the two increases .

Thus χ^2 affords a measure of correspondence b/w theory & Observation.

Note : If the data is given in a series of 'n' no.s then $d.f = n-1$.

In case of Binomial Distribution : $d.f = n-1$

" " " Poisson " : $d.f = n-2$

" " " Normal " : $d.f = n-3$.

If O_i $(i=1,2,\ldots n)$ is set of Observed (experimental) frequencies, &

E_i $(i=1,2,\ldots n)$ is corresponding set of expected (theoretical) frequencies, then

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{O_i - E_i}{E_i}\right)^2$$

$$\boxed{\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}}$$

with $\nu = n-1$ d.f .

χ^2 Test as Test of Goodness of Fit

Conditions of Validity:

The following conditions should be satisfied before χ^2 Test is applies.

(i) Sample Observations should be independent.

(ii) Total frequency (ie N) is large. ie $N > 50$.

(iii) The constraints of on the cell frequencies, if any, are linear.

(iv) No theoretical (or expected) frequency should be less than 10.

If they occur, then the difficulty is overcome by grouping 2 or more classes together before finding $(O-E)$. Also the d.f is determined with no. of classes after regrouping.

H_0: There is no significant difference b/w Observed values & expected values.

H_1: The difference is significant.

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

$\nu = n-1$ d.f.

α = L.O.S.

If Calculated $\chi^2 <$ Tabulated $\chi^2 \Rightarrow$ Accept H_0

Otherwise Reject H_0.

Do problems

Chi-Square Test for Independence Of Attributes :

"Attribute" literally means a quality or characteristic.
eg: drinking, smoking, blindness, honesty, beauty etc.
An Attribute may be marked by its presence (position) or absence in a no. of a given population.

Let Observed frequencies be classified according to two attributes say A, B & the frequencies O_i in the different categories be shown in a two-way table called contingency table.

On the basis of cell frequencies we have to Test whether the 2 attributes are independent or not.

The Expected frequencies (E_i) of any cell is

$$E_i = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}.$$

ie. If

A	a	b	a+b
B	c	d	c+d
	a+c	b+d	N=a+b+c+d

Its Expected frequencies are:

$$E(a) = \frac{(a+c)(a+b)}{N\,(=a+b+c+d)}$$

$$E(b) = \frac{(b+d)(a+b)}{N};\ E(c) = \frac{(a+c)(c+d)}{N};\ E(d) = \frac{(b+d)(c+d)}{N}$$

The table is

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
a	E(a)	a − E(a)	- - -	- - -
b	E(b)	b − E(b)	- - -	- - -
c	E(c)	c − E(c)		
d	E(d)	d − E(d)		

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i-E_i)^2}{E_i}$$

Degrees of freedom: $\nu = (\text{No. of rows} - 1) \times (\text{No. of Columns} - 1)$.

$\Rightarrow \nu = (r-1)(c-1)$.

Unit-IV

Problems on Exact Sampling Distributions (Small Samples)

①

1. A sample of 26 bulbs gives a mean life of 990 hours with a S.D of 20 hrs. The manufacturer claims that mean life of bulbs is 1000 hrs. Is the sample not up to the Standard.

Sol. Given $n=26$; $\bar{x}=990$, $s=20$

Let Null hypothesis H_0: $\mu=1000$ hrs.

Alternative Hypothesis H_1: $\mu<1000$

Level of significance: $\alpha=0.05$ & d.o.f $(n-1)$

Test statistic

$$t_{cal}=\frac{\bar{x}-\mu}{s/\sqrt{n-1}}=\frac{990-1000}{20/\sqrt{25}}=\frac{-10}{4}=-2.5$$

But t_{tab} at $\alpha=0.05$ with 25 d.o.f $=1.708$

$\because |t_{cal}|>t_{tab}$; we reject the Null Hypothesis H_0

i.e $\mu\neq 1000 \Rightarrow$ The manufacturer claim Mean of bulbs 1000 hrs is not Correct.

i.e the sample is not up to the Standard.

2. A sample of 25 members has a mean 67 and S.D 5.2 Is this sample has been taken from a large population of mean 70?

Sol. Given $n=25$, $\bar{x}=67$; $s=5.2$

Let Null Hypothesis H_0: $\mu=70$

Alternative Hypothesis H_1: $\mu\neq 70$

Level of Significance : $\alpha = 0.05$ with ~~2~~4 d.o.f

Test statistic

$$t_{cal} = \frac{\bar{x}-\mu}{s/\sqrt{n-1}} = \frac{67-70}{5.4/\sqrt{\cancel{25}5-1}} = \frac{-3}{1.6614} = 2.82.$$

$$\therefore |t_{cal}| = 2.82 > 2.06$$

we reject the Null hypothesis and concluding that the sample has not taken from population of mean 70.

3. 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 means I.Q values of samples of 10 boys?

Sol^n Given $n = 10$
Here S.D and mean of the sample are not given we have to find Mean & S.D of sample

$$\text{So, Mean} = \frac{70+120+110+101+88+83+95+98+107+100}{10} = \frac{972}{10} = 97.2$$

$$\therefore \bar{x} = 97.2$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{10}(x-\bar{x})^2 = \frac{1}{9}\left[(70-97.2)^2+(120-97.2)^2+\cdots+(100-97.2)^2\right]$$

$$= \frac{1833.6}{9} = 203.73$$

$$\Rightarrow s = \sqrt{203.73} = 14.27.$$

Let Null Hypothesis $H_0 : \mu = 100$

Alternative Hypothesis $H_1 : \mu \neq 100$ (Two tailed test)

Level of significance : $\alpha = 0.05$ with 9 d.o.f

Test Statistic

$$t_{cal} = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{97.2-100}{14.27/\sqrt{10}} = -0.62$$

$\therefore |t_{cal}| = 0.62$

Since $|t_{tab}|$ with 9 d.o.f and $\alpha = 0.05$ is 2.262

we accept Null Hypothesis.

$\therefore$ the assumption of mean I.Q of 100 in the population is correct. (supportive)

b) The 95% confidence limits $= \left(\bar{x} - t_{0.05 \text{ with } 9 \text{ d.o.f}} \cdot (S/\sqrt{n}),\ \bar{x} + t_{0.05}(S/\sqrt{n})\right] = \left(97 \pm 2.26\left(\frac{14.22}{\sqrt{10}}\right)\right) = 87$ and 107.4.

4. 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 average height is greater than 64 inches? Test at 1% level of significance

Sol. Given $n = 10$; Mean $= \frac{\Sigma x}{n} = \bar{x}$; $S^2 = \frac{1}{n-1}\sum_1 (x-\bar{x})^2$

x	$x-\bar{x}$	$(x-\bar{x})^2$
70	4	16
67	1	1
62	-4	16
68	2	4
61	-5	25
68	2	4
70	4	16
64	-2	4
64	-2	4
66	0	0
660 Σx		90 $\Sigma(x-\bar{x})^2$

$\therefore \bar{x} = 66$

and $S^2 = \frac{1}{9}[90] = 10$

$S = \sqrt{10} = 3.16$

1. Null Hypothesis H_0: $\mu = 64$
2. Alternative Hypothesis H_1: $\mu > 64$
3. Level of significance: $\alpha = 0.01$ with 9 d.o.f
4. Test statistic

$$t_{cal} = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{66-64}{3.16/\sqrt{10}} = \frac{2\sqrt{10}}{3.16} = 2.0014$$

Here t_{tab} at 0.01 level of significance with 9 d.o.f $= 2.821$

Since $t_{cal} < t_{tab}$ we accept null Hypothesis

$\therefore$ At 1% level of significance $\mu = 64$.

5. Two horses A and B were tested according to the time (in sec) 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 $n_1 = 7$; $n_2 = 6$

Degrees of freedom $= n_1 + n_2 - 2 = 7+6-2 = 11$ d.o.f

Mean of series A $= \bar{x} = \frac{\Sigma x}{n_1} = \frac{28+30+32+66+29+34}{7} = 31.286$

Mean of series B $= \bar{y} = \frac{\Sigma y}{n_2} = \frac{29+60+24+27+29}{6} = 28.16$

Now $S^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2\right]$

x	$(x-\bar{x})$ $x-31.286$	$(x-\bar{x})^2$	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	3.2856
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

$\therefore \Sigma(x-\bar{x})^2 = 31.4358$ and $\Sigma(y-\bar{y})^2 = 26.8336$

$\therefore S'^2 = \frac{1}{11}[31.4358 + 26.8336] = \frac{58.2694}{11} = 5.23$

$\therefore S = \sqrt{5.23} = 2.3$

1. Null Hypothesis H_0: $\mu_1 = \mu_2$
2. Alternative Hypothesis H_1: $\mu_1 \neq \mu_2$
3. Level of significance : $\alpha = 0.05$ with 11 d.o.f
4. Test statistic

$$t_{cal} = \frac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{31.286 - 28.16}{2.3\sqrt{\frac{1}{7}+\frac{1}{6}}} = 2.443$$

Here $t_{tab} = 2.201$

Since $t_{cal} > t_{tab}$, we reject the Null Hyp H_0

$\Rightarrow$ There is the Significant difference in the running capacity of Horses A and B.

6. find the Maximum difference that we can expect with probability 0.95 b/w the means of samples of sizes 10 and 12 from a normal population if their S.DS are found to be 2 and 3 respectively.

Soln Given $n_1 = 10$ and $n_2 = 12$

$s_1 = 2$, $s_2 = 3$

then $S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{10(4) + 12(9)}{20} = \frac{40 + 108}{20} = \frac{\cancel{148}^{74}}{\cancel{20}_{10}} = 7.4$

$\therefore S = \sqrt{7.4} = 2.72.$

Test statistic

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Here t_{tab} with 95% confinence $= 2.086$

[$\because \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$

$t_{0.025}$ with 20 d.o.f $= 2.086$]

$$\therefore (2.086) = \frac{\bar{x} - \bar{y}}{2.72\sqrt{\frac{1}{10} + \frac{1}{12}}}$$

$$\Rightarrow \bar{x} - \bar{y} = (2.086)(2.72)\sqrt{\frac{6+5}{60}}$$

$$\bar{x} - \bar{y} = 2.43.$$

$\therefore$ Maximum difference between the means of Samples $= 2.43$

7. The Blood pressure of 5 woman before and after intake of a certain drug are given below.

Before	110	120	123	132	125
after	120	118	125	136	121

Test whether there is Significant change in Blood pressure at 1% level of significance.

Sol. Let Null Hypothesis $H_0: \mu_1 = \mu_2$

Alternative Hypothesis $H_1: \mu_1 < \mu_2$

Level of Significance: $\alpha = 0.01$ with 4 d.o.f $\underline{(n-1)}$

Test Statistic

$$t = \frac{\bar{d}}{s/\sqrt{n}} \quad \text{where } \bar{d} = \frac{\Sigma d}{n} \text{ and } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d-\bar{d})^2$$

$$= \frac{-12}{5} = -2.4$$

x (BP before intake of drug)	y BP after intake	d = x−y	$d-\bar{d}$ $d+2.4$	$(d-\bar{d})^2$
110	120	−10	−7.6	57.76
120	118	2	4.4	19.36
123	125	−2	0.4	0.16
132	136	−4	−1.6	2.56
125	121	4	6.4	40.96
		$\Sigma d = -12$		$\Sigma(d-\bar{d})^2 = 120.8$

$$\therefore s^2 = \frac{1}{4}(120.8) = 30.2$$

$$\Rightarrow s = \sqrt{30.2} = 5.4954$$

$$\therefore t_{cal} = \frac{-2.4}{5.4954/\sqrt{5}} = \frac{-2.4\sqrt{5}}{5.4954} = \frac{-5.3666}{5.4954} = -0.9765$$

But t_{tab} at $\alpha = 0.01$ with d.o.f $= 4$ is 4.604

Since $|t_{cal}| < t_{tab}$, we accept H_0

$\therefore$ There is no significant change in BP after intake of certain Drug.

8. The average losses of workers, before and after certain program are given below. Use 0.05 level to test whether the program is effective (4)

Before	40	70	45	120	35	55	77
After	35	65	42	116	33	50	73

Sol: Null hypothesis $H_0: \mu_1 = \mu_2$

Alternative hypothesis $H_1: \mu_1 \neq \mu_2$

level of Significance: $\alpha = 0.05$ with '6' d.o.f

Test statistic

$$t_{cal} = \frac{\bar{d}}{S/\sqrt{n}} \quad \text{where } \bar{d} = \frac{\Sigma d}{n} = \frac{28}{7} = 4$$

$$\text{and } S^2 = \frac{1}{n-1}\Sigma(d-\bar{d})^2$$

Before (x)	After (y)	$d = x - y$	$d-\bar{d}$	$(d-\bar{d})^2$
40	35	5	1	1
70	65	5	1	1
45	42	3	-1	1
120	116	4	0	0
35	33	2	-2	4
55	50	5	1	1
77	73	4	0	0
		28		8

$$\therefore S^2 = \frac{1}{\cancel{6}_3}(\cancel{8}^4) = \frac{4}{3} = 1.333$$

$$S = \sqrt{1.333} = 1.1545$$

$$\therefore t_{cal} = \frac{4}{1.1545/\sqrt{7}} = \frac{4\sqrt{7}}{1.1545} = 9.167$$

Here $t_{tab} = 2.447$

$\because t_{cal} > t_{tab}$; we reject H_0

$\therefore$ The program is not effective

9. Memory capacity of 10 students were tested before and after training. State whether the training was effective or not from the following scores.

Before	12	14	11	8	7	10	3	0	5	6
After	15	16	10	7	5	12	10	2	3	8

Sol. Solve as previous problems.

[Hint: t_{tab} with $\alpha = 0.05$, $d.o.f = 10-1 = 9$ is 1.833

$H_0: \mu_1 = \mu_2$ $\quad t_{cal} = -1.365$

$H_1: \mu_1 < \mu_2$ $\quad |t_{cal}| < t_{tab} \Rightarrow$ ~~Reject~~ Accept H_0

$\therefore$ Training was not effective.

$\Rightarrow$ There is no significant difference b/w before and after training in scores.

10. 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

Sol. To test whether the two samples came from same population, we have to test

i) the equality of means by using student 't' test

ii) the equality of variances by using F-test

Sol. Given $n_1 = 5$; $n_2 = 6$

Mean and S.D's of the Samples

x	$x-\bar{x}$ $x-24.6$	$(x-\bar{x})^2$	y	$y-\bar{y}$ $y-29$	$(y-\bar{y})^2$
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

$\therefore \bar{x} = \frac{\Sigma x}{n_1} = \frac{123}{5} = 24.6$

$\bar{y} = \frac{\Sigma y}{n_2} = \frac{174}{6} = 29$

$\Sigma(x-\bar{x})^2 = 21.2$ and $\Sigma(y-\bar{y})^2 = 108$

$S_1^2 = \frac{1}{n_1-1}\Sigma(x-\bar{x})^2$

$= \frac{21.2}{4} = 5.3$

$S_2^2 = \frac{1}{n_2-1}\Sigma(y-\bar{y})^2$

$= \frac{108}{5}$

$= 21.6$

(i) F-Test:

Null Hypothesis H_0: $\sigma_1^2 = \sigma_2^2$

Alternative Hyp H_1: $\sigma_1^2 \neq \sigma_2^2$

Level of significance: $\alpha = 0.05$ with $(5,4)$ d.o.f

Test statistic

$F_{cal} = \frac{S_2^2}{S_1^2}$ ($\because S_2 > S_1$).

$= \frac{21.6}{5.3} = 4.075$

But $F_{tab} = 6.26$; $\because F_{cal} < F_{tab}$, we accept H_0

$\therefore$ The population (normal) have equal variances

(ii) Student's t test:

Null hyp $H_0: \mu_1 = \mu_2$

Alternative hyp $H_1: \mu_1 \neq \mu_2$

Level of significance: $\alpha = 0.05$ with 9 d.o.f

Test Statistic

$$S^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2\right]$$

$$= \frac{1}{9}[21.2+108]$$

$$= \frac{129.2}{9} = 14.35$$

$$\therefore S = \sqrt{14.35} = 3.78$$

Now $t_{cal} = \dfrac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \dfrac{24.6-29}{3.78\sqrt{\frac{1}{5}+\frac{1}{6}}} = -1.92$

But $t_{tab} = 2.262$

$\because |t_{cal}| < t_{tab}$ we accept H_0

$\therefore$ The normal population come from equal means.

11. Pumpkins were grown under two experimental Condns Two random samples of 11 and 10, show the sample S.D of their weights as 0.8 and 0.5 respy. Assume that the weight distribution are normal, Test the hypothesis is that the true Variances are equal.

Sol. Given $n_1 = 11$ and $n_2 = 10$

$S_1 = 0.8 \qquad S_2 = 0.5$

$$\Rightarrow S_1^2 = \frac{1}{n_1-1}\Sigma(x-\bar{x})^2 = \frac{n_1 s_1^2}{n_1-1} = \frac{11}{10}(0.8)^2 = 0.704 \Rightarrow S_1 = \sqrt{\frac{11(0.8)^2}{10}} = 0.839$$

$$s_2'^2 = \frac{1}{n_2-1}\Sigma(y-\bar{y})^2 = \frac{n_2 s_1^2}{n_2-1} = \frac{10}{9}(0.5)^2 = \cancel{0.2685}$$

$$\Rightarrow s_2' = \sqrt{\frac{10}{9}}(0.5) = \cancel{0.2685}\ 0.52704$$

$\because s_2'^2 > s_1^2$

[Here Null hyp H_0: $\sigma_1^2 = \sigma_2^2$
Alternative hyp H_1: $\sigma_1^2 \neq \sigma_2^2$
Level of significance: $\alpha = 0.05$

Test statistic

$$F_{cal} = \frac{s_2^2}{s_1^2} = \frac{0.2704}{0.2777} = 2.535$$ with (n_1-1, n_2-1) d.o.f]

But $F_{tab} = 3.14$

$\because F_{cal} < F_{tab}$, we accept H_0

$\therefore$ True Variances are equal.

12 Four coins were tossed 160 times and the following results were obtained.

No of heads	0	1	2	3	4
observed freq	17	52	54	31	6

Under the assumption that coins are balanced, find the Expected frequencies of 0, 1, 2, 3, 4 heads and Test the goodness of fit.

<u>Sol</u>

Sol^n Here probability of getting a head in a single throw $= p = \frac{1}{2} \Rightarrow q = 1 - \frac{1}{2} = \frac{1}{2}$ Prob of success = getting head

Probability of getting '0' heads $= 4C_0\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4 = \frac{1}{16}$

$\therefore$ Expected frequency for '0' heads

$= E(0) = \frac{1}{16} \times 160 = 10$ [$\because N = 160$ times
$n = 4$
$P(X=r) = nC_r\, p^r q^{n-r}$]

13. 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 cond^ly were the same during 10 week period.

Sol^n Let the Null Hyp H_0: The accident cond^ly were same during 10 week period

Alternative Hyp H_1: Accident cond^ly were not same

Level of significance: $\alpha = 0.05$ with 9 d.o.f.

O_i	E_i	$O_i - E_i$	$(O_i-E_i)^2$	$(O_i-E_i)^2/E_i$
12	10	2	4	0.4
8	10	−2	4	0.4
20	10	10	100	10
2	10	−8	64	6.4
14	10	4	16	1.6
10	10	0	0	0
15	10	5	25	2.5
6	10	−4	16	1.6
9	10	−1	1	0.1
4	10	−6	36	3.6
100	100			26.6

Test statistic

$$\chi^2_{cal} = \sum \frac{(O_i - E_i)^2}{E_i} = 26.6$$

$\because \chi^2_{tab} = 16.916$, we reject Null Hypothesis

$\therefore$ the accident cond^ly were not same during the 10 week period.

$E(1) = N\, {}^4C_1 \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 = N\, 4\, \frac{1}{16} = \frac{1}{4}(\overset{40}{160}) = 40$

$E(2) = N\, {}^4C_2 \left(\frac{1}{2}\right)^4 = \frac{\overset{10}{160}}{16}\, \frac{4!}{2!\,2!} = \frac{10 \times \overset{2}{4} \times 3}{2} = 60$

$E(3) = N\, {}^4C_3 \left(\frac{1}{2}\right)^4 = \frac{10 \times 4!}{3!} = 10 \times 4 = 40$

$E(4) = N\, {}^4C_4 \left(\frac{1}{2}\right)^4 = 160 \cdot \frac{1}{16} = 10$

$\therefore$ The table is

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
17	10	7	49	4.9
52	40	12	144	3.6
54	60	-6	36	0.6
31	40	-9	81	2.025
6	10	-4	16	1.6
160 ΣO_i	160 ΣE_i			12.725

Let Null Hyp H_0: The data follows Binomial distribution

Alternative Hyp H_1: The data not follows B.D

level of significance: $\alpha = 0.05$ with 4 d.o.f

[one d.o.f is lost for parameter μ]

Test statistic

$$\chi^2_{cal} = \Sigma \frac{(O_i - E_i)^2}{E_i} = 12.725$$

$\therefore \chi^2_{tab}$ at $\alpha = 0.05$ with 4 d.o.f $= 9.488$

we reject H_0 [$\because \chi^2_{tab} < \chi^2_{cal}$]

$\therefore$ The data not follows Binomial Distribution

UNIT-V
Queueing Theory

Queue:- The group of items waiting for the service including the customer, who is receiving the service is called a Queue.

The places where queue is needed are Bank, post-office, Railway station, Airport, Bustop, etc.,

~~customer~~

Queueing theory:-

Queueing is an activity in which every individual may experience in their daily lives. It involves two parties.

(i) customers (ii) Server.

Customers:- The person who is waiting for the service is the customer.

Server :- The person who is servicing the customers is a Server.

Describe a Queueing System

A Queueing system can be completely described by the following components

1) **Input pattern (or) Arrival pattern:-**

The input describes the way in which the customers arrive and join the system. The ability of Queueing system to provide service for an arriving group of customers depends not only the mean arrival rate, but also on the pattern in which they arrive.

2) <u>Service pattern</u>:-

The way in which the server serves is known as the service pattern. It is possible that there may be single (one) server (or) several (may) servers to serve the customers in Queue. The time taken to serve a customer by the server is referred to as "service time", which is usually a random variable.

The service time distribution is taken as "negative exponential distribution.

3) <u>Queue discipline</u>:-

It is a rule according to which customers are selected for service when Queue has been formed. The most common Queue discipline is the "First come, First served" (FCFS) (or) the "First in, First out" (FIFO) rule under which the customers are served in the strict order of their arrivals. other queue discipline include "Last in, First out (LIFO) rule according to which the last arrival in the system is serviced first.

4) <u>Queue Behaviour</u>:- The customers generally behave in four ways

balking :- customers may not enter the Queue in view of its length. This customer behaviour is referred to as "balking".

Reneging :- This occurs when a waiting customer leaves the Queue due to impatience (or) importance.

Priorities :- Some customers may be served according to the priority without waiting in the Queue.

Jockeying :- If a customer Jump from one Queue to another, if there are more than one queue, then it is Jockeying.

Operating characteristics of a Queueing system.

Some of the operational characteristics of a Queueing system, that are of general interest for the evalusion of the performance of an existing Queuing system and to design a new system are as follows.

i) **Expected no. of customers in the system** :-

It is denoted by E_n (or) L_s is average no. of customers in the system, both waiting and in service. Here n = no. of customers in the Queueing system

2) Expected no. of customers in the Queue:-

It is denoted by E_m (or) L_q is the Average no. of customers waiting in the Queue. Here $m = n-1$ i.e excluding the customer being served.

3) Expected waiting time in the system:-

It is denoted by E_v (or) W_s is the Average total time spent by a customer in the system. It is generally taken to be the waiting time plus servicing time.

4) Expected waiting time in Queue:-

It is denoted by E_w (or) W_q is the Average time spent by a customer in the Queue before the commencement of his service.

5) The Server utilization factor (or) busy period:-

It is denoted by $\rho\ (= \frac{\lambda}{\mu})$ is the proportion of time that a server actually spends with the customer. Here λ = the Average no. of customers arriving per unit of time and μ = the Average no. of customers completing service per unit of time.

The server utilization factor is also known as "traffic intensity (or) the clearing ratio.

Definations :-

Queue length :-

The no. of customers waiting in the Queue at any time is called the Queue length.

Average Queue length :-

No. of customers in the Queue per unit time.

Waiting time :-

The time a customer (unit) has to wait in the Queue till he (it) will be taken into Service.

Busy period :-

The time when the server is busy in serving. So this is the time from the starting of the service unit to the last unit.

Servicing time :-

The time taken to serve a unit is the servicing time.

Idle time :-

The time when the server is not servicing any customer (unit)

Mean arrival rate :-

Average no. of arrivals in a time interval of unit length.

Mean servicing rate :-

The Average no. of units served in a time interval of unity.

Traffic Intensity :-

The ratio of the mean arrival rate to the mean service rate

$$\rho = \frac{\lambda}{\mu}$$

Transiant State and Steady States

Transiant state :-

If the operating characteristics are depends on time. Then their system said to be in transient State.

So in this state waiting time, Service time the probability distribution are dependent on time, the System is in transiant.

Steady State :-

If the Operating characteristics are independent of time. Then the System said to be steady State.

So In this system waiting time, Serverice time distribution of arrivals are independent of time.

Explosive State :- If the arrival rate is greater than the service rate. Then the system cannot attain the steady state. The Queue length increases rapidly with time and tend to ∞. This state is called Explosive State.

Notations & Symbols of Queueing system.

n - no. of Customers (units) in the system including one that is served.

m - no. of customers in the Queue Excluding the one that is served.

$P_n(t)$ = The transident probability that these are n - customers in the system at any time "t".

P_n - The steady state prophability that there are n - customers in the system at any time "t".

λ = no. of arrivals per unit time.

μ = no. of services per unit time.

ρ = traffic Intensity $= \left(\frac{\lambda}{\mu}\right)$

L_s = Expected (Average) no. of customer in the system.

L_q = Expected (Average) no. of customer in the Queue.

Probability distribution in Queueing Systems.

It is assumed that Customers Joining the Queueing system arrive in a random manner and follow a poisson distribution (or) equivalently the inter-arrival times obey Exponential distributed.

The arrival and service distributions for poisson Queues are derived. The basic assumptions (Axioms) governing this type of Queues are stated below.

Axiom 1:- The no. of arrivals in non-over lapping intervals is statistically independent.

i.e The process has independent increments

Axiom 2:- The probability of more than one arrival b/w time t & $t+\Delta t$ is $o(\Delta t)$

Axiom 3:- The probability that an arrival occurs b/w time t & $t+\Delta t$ is equal to $\lambda \Delta t + o(\Delta t)$ where λ is a Constant. Δt is an incremental element and $o(\Delta t)$ represents the terms such that

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$$

A process characterised by the above axioms is called a poisson process. This is equivalent to Assuming that the inter arrival time is an exponentially distributed random variable.

→ If in the above λ is replaced by μ. The no. of serivces per unit time, we note that the service distribution also is a poisson process with inter-service time following an Exponetial distribution.

Arival distribution theorem

If the arrivals are completely random. then the probability distribution of no. of arrivals in a fixed time-interval follows a poisson distribution.

Kendall's Notation

The representation of Queueing model, as given below was introduced by D.G. Kendall and then by A. Lee

$$(a/b/c) : (d/e)$$

a – arrival distribution, b – Departure distribution

c – No. of Channels, d – maximum. no of customers allowed

e – Queue disciple. i.e capacity of the system.

a & b are represented by M & M

M = Markovian (poisson arrival & Exponential distribution)

Pure Birth and Death process.

Distribution of arrivals (or) pure Birth process :-

A Queueing model in which only arrivals are counded and no departures take place are called pure birth process.

Distribution of Departures (or) pure Death process :-

A Queueing model in which only departures are counded and no other arrivals allowed are called pure death process.

pure Birth & Death process :-

In Queueing model, we have arrivals (or births) & departures (or deaths). Such a model (or) process is called a pure Birth & death model (or) process.

Classification of Queuing models.

The Queueing models can be categorized as:-

1) Deterministic Queueing model :-

If each customer arrives at known intervals and the services time is known with certainty. The Queueing model is said to be deterministic in nature.

2) Probabilistic Queueing model :-

If the arrivals of the customers (or) the services times of the customers (or) both of Queueing system is not known with certainty & expressed only in probabilistic nature, then the Queueing model is said to be probabilistic nature.

i) Model - I

(M/M/1) : (∞/FCFS).

The first M - denotes that the arrivals are poisson.

The second M - denotes that the departures are poisson

1 - denotes that single service channel.

∞ - denotes that the arrivals are from an infinite population and there is no limit on the system capacity.

FIFO : First in, first out.

Characteristics of (M/M/1) : (∞/FIFO) system

1. λ = no. of arrivals per unit time.

2. μ = no. of services per unit time.

3. ρ = traffic intensity = $\frac{\lambda}{\mu}$

4. The probability that the system is busy = ρ

5. P_n = probability that there 'n' people in the system
$= \rho^n(1-\rho)$

6. P_0 = The probability that the system is idle $= 1-\rho$

7. P (There are K (or) more customers in the system)

$$= P(n \geq K) = \rho^K.$$

8. The probability that the no. of customers in the system exceeds K (or) greater than $K = P(n > K) = \rho^{K+1}$

9. Average no. of customers in the system (or) Average length of the System $= L_s = \dfrac{\rho}{1-\rho}$

10. Average no. of customers in the Queue (or) Average length of the Queue $= L_q = \dfrac{\rho^2}{1-\rho}$

11. Average no. of customers in the non-empty Queue

$$= E(m/m>0) = \frac{\mu}{\mu-\lambda}$$

12. Variance of Queue Length $= \dfrac{\rho}{(1-\rho)^2}$

13. The probability density function (Cummulative probability distribution) of waiting time distribution (exculding service time) $= \psi(\omega) = \lambda\left(1-\frac{\lambda}{\mu}\right)e^{-(\mu-\lambda)\omega}$ if $\omega > 0$

14. The probability that a customer has to wait for more than t minutes (probability that waiting time exceeds t)

$$= \int_t^\infty \frac{\lambda}{\mu}(\mu-\lambda) \cdot e^{-(\mu-\lambda)t}\, dt = \frac{\lambda}{\mu} \cdot e^{-(\mu-\lambda)t}$$

15. Average waiting time of a customer in the Queue $= W_q$

$$W_q = \frac{\rho}{\mu - \lambda}$$

16. Average waiting time of a customer in a non-empty

Queue $= E(w / w > 0) = \frac{1}{\mu - \lambda}$

17. Average time that a customer spends in the system

$$W_s = \frac{1}{\mu - \lambda}$$

18) Litle's formulae

$$L_s = \lambda W_s, \quad L_q = \lambda W_q$$

$$L_q = \rho L_s, \quad W_q = \rho W_s$$

STOCHASTIC PROCESSES

Defination :- A Stochastic process is a set of random variable $\{x_i\}$ (or) $\{x_t\}$ depending on same real parameters like time 't'.

→ **States :-** The values assumed by the random variable $x(t)$ are called states.

State space :-

The set of all possible values of any individual member of the random process is called "state space". It is denoted by I (or) S.

→ The state space is said to be discrete if it contains a finite (or) countable infinity of points otherwise it is called continuous.

→ If the parameter set T is Discrete. The Random process will be denoted by $\{x_n\}$ (or) $\{x(n)\}$

→ If the parameter set T is continous the process will be denoted by $\{x(t)\}$ (or) $\{x_t\}$.

Examples of Stochastic process

1) A Queuing System
2) Turblent Fluid Flow
3) Movement of mole cules of a gas (or) liquid.

4) A Random walk model.

5) communication process.

6) Gamblers Ruin problem.

Classification of Stochastic processes

Stochastic processes mainly classified into four types. Based on time 't' and the random variable x. They are:-

1) **Continuous stochastic processes:-**

If both the random variable x and time t are continuous, the stochastic processes is called a "continuous Stochastic process.

2) **Discrete Stochastic processes:-**

If the random variable x is discrete & time 't' is continuous, The stochastic process is called a "Discrete Stochastic process".

3) **Discrete Stochastic Sequence:-**

If both random variable x & time 't' are discrete. Then the stochastic process is called a "Discrete Stochastic Sequence".

4) **Continuous Stochastic Sequence:-**

If the random variable x & is continuous and time 't' is discrete then the stochastic process is called "Continuous random Sequences".

Deterministic Stochastic process.

A random process is called a Deterministic stochastic process if future values of any sample function can be predicted from its part observations (or) part values.

Non - Deterministic stochastic process.

A Stochastic process is called non- Deterministic Stochastic process if future values of any sample function cannot be predicted from part observations (or) part values.

Stochastic process with Independent Increments.

If, for all $t_1 < t_2 < \dots < t_n$, The random variables $x(t_2) - x(t_1)$, $x(t_3) - x(t_2)$, ..., $x(t_n) - x(t_{n-1})$ are Independent. Then the process $\{x(t)\}$, is said to be a stochastic process with independent increments.

Methods of description of random process

Consider a random process $x(t)$. For any given time t_1, the distribution associated with a Random process x_1 is given by $F(x_1, t_1) = P(x(t_1) < x_1)$. is called First order distribution of the random variable $x(t_1)$ and $f(x_1, t_1) = \frac{\partial}{\partial x_1} F(x_1, t_1)$ is called First order probability density function of random variable.

$\{x(t)\}$ Illy $F(x_1, x_2, t_1, t_2) = P(x(t_1) \le x_1, x(t_2) \le x_2\}$ is the joint probability distribution of the Random Variables $x(t_1)$ & $x(t_2)$ is called the second order distribution of the random variable & $f(x_1, x_2, t_1, t_2) = \frac{\partial^2}{\partial x_1, \partial x_2} F(x_1, x_2, t_1, t_2)$ is called the Second order density function of random variable $x(t)$

Illy the n^{th} order distribution $\{x(t)\}$ is the joint distribution $F(x_1, x_2 \ldots x_n); t_1 t_2 \ldots t_n\}$ of the Random variables $x(t_1), x(t_2) \ldots x(t_n)$

Stationary process and non- stationary process

1) Stationary process :-

If certain probability distribution (or) average donot depend on time 't'. Then the random process $\{x(t)\}$ is called stationary process.

Depending on the density functions of the random variables of the process, there are several types of stationary they are.

i) First order stationary process :-

If the first order probability density function of random process is independent of time, then it is called as First order stationary process.

$F(x_1, t_1) = F(x_1, t_1+\delta)$ where δ is a real number.

Stationary process:-

(ii) Second order :- If the second order probability density function of random process is independent of time, Then it is called as second order stationary process.

$$f(x_1, x_2, t_1, t_2) = f(x_1, x_2; t_1+\delta, t_2+\delta).$$

where δ is real number

lly if the order probability density function of random process is independent of time. Then it is called as n^{th} order stationary process.

$$f(x_1, x_2 \dots x_n, t_1, t_2, \dots t_n) = f(x_1, \dots x_n, t_1+\delta, \dots t_n+\delta).$$

where δ is real number.

2) Non-Stationary process:-

A process which is not stationary is said to be non-stationary (or) evolutionary process.

Time Averages:-

Consider stochastic (or) random process $x(t)$. The Average value of $x(t)$ taken over all times is called the time average of $x(t)$. It can be expressed as

$$\bar{x} = A\left[x(t)\right] = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x(t)\, dt$$

Average Values of Single Random process and two or more random process

The Average values of Single random process (or) two (or) more random process are as follows.

1) **Mean :-**

Mean function of random process $x(t)$ is expected value of a number of random process.

i.e. Mean $\mu(t) = E(x(t))$

2) **Auto Correlation :-**

The Auto correlation of the random process $x(t)$ is the expected value of the product of any two random variables $x(t_1)$ & $x(t_2)$ of the process and It is denoted $R(t_1, t_2)$ (or) $R_x(t_1, t_2)$ (or) $R_{xx}(t_1, t_2)$

$$\therefore R(t_1, t_2) = E\left[x(t_1) \cdot x(t_2)\right]$$

3) **Auto Covariance :-**

The Auto - Covariance of the random process $x(t)$ is denoted by $C(t_1, t_2)$ (or) $C_x(t_1, t_2)$ (or) $C_{xx}(t_1, t_2)$ and defined as the covariance of the random variables

$x(t_1)$ & $x(t_2)$.

$$\therefore c(t_1, t_2) = E\left[\{(x(t_1) - \mu(t_1)\}\{x(t_2) - \mu(t_2)\}\right]$$

$$c(t_1, t_2) = R(t_1, t_2) - \mu(t_1) \cdot \mu(t_2).$$

4) <u>Correlation Co-efficient :-</u>

The correlation co-efficient of random process $x(t)$ is defined as

$$\rho(t_1, t_2) = \frac{c(t_1, t_2)}{\sqrt{c(t_1, t_1) \cdot c(t_2, t_2)}}$$

Where $c(t_1, t_1)$ = Variance of $x(t_1)$ &

$c(t_2, t_2)$ = Variance of $x(t_2)$

5) <u>Cross - Correlation :-</u>

The cross correlation of the random process $x(t)$ & $x(t)$ is defined as

$$R_{xy}(t_1, t_2) = E\{x(t_1) \cdot y(t_2)\}$$

6) <u>Cross - Covariance :-</u>

The cross covariance of the random process $x(t)$ & $y(t)$ is defined as the covariance of two random variables $x(t_1)$ & $y(t_2)$.

$$\therefore C_{xy}(t_1, t_2) = R_{xy}(t_1, t_2) - \mu_x(t_1) \cdot \mu_y(t_2)$$

7) Cross - correlation Co-efficient.

The cross - correlation co-efficient of two random process $x(t)$ & $Y(t)$ is defined as

$$\rho_{xy}(t_1, t_2) = \frac{C_{xy}(t_1, t_2)}{\sqrt{C_{xx}(t_1, t_1) \cdot C_{yy}(t_2, t_2)}}$$

$$\rho_{xy}(t_1, t_2) = \frac{C_{xy}(t_1, t_2)}{\sqrt{C_{xx}(\,,\,) \cdot C_{y}}}$$

* If $p = q$ then (i) probability of ruin $q_z = \frac{a-z}{a}$

(ii) Expected duration of the game $d_z = z(a-z)$

* If $p \neq q$ Then (i) probability of ruin $= q_z = \dfrac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^z}{\left(\frac{q}{p}\right)^a - 1}$

(ii) Expected duration of the game

$$d_z = \left(\frac{-a}{q-p}\right) \frac{\left(\frac{q}{z}\right)^z - 1}{\left(\frac{q}{p}\right)^a - 1} + \frac{z}{q-p}$$

1) Calculate the probability of ruin and expected duration of the game, where

(i) $a = 50$, $z = 40$, $p = 0.5$

(ii) $a = 100$, $z = 5$, $p = 0.6$

<u>Sol</u> (i) Given that $a = 50$, $z = 40$, $p = 0.5 = \frac{1}{2}$

$$\therefore q = 1 - 0.5$$
$$= 0.5$$
$$= 1/2$$

$\therefore$ probability of ruin $= q_z = \frac{a-z}{a} = \frac{50-40}{50} = \frac{10}{50} = 0.2$

Expected duration of game $= d_z = z(a-z)$

$$= 40(50-40)$$
$$= 40 \times 100 = 400$$

(ii) Given that $a = 100$; $z = 5$, $p = 0.6$

$$Q = 1 - 0.6$$
$$= 0.4$$

$$\therefore \frac{p}{q} = \frac{2}{3}$$

probability of ruin $= q_z = \dfrac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^z}{\left(\frac{q}{p}\right)^a - 1}$

$$q_z = \frac{\left(\frac{2}{3}\right)^{100} - \left(\frac{2}{3}\right)^5}{\left(\frac{2}{3}\right)^{100} - 1} = 0.132$$

expected duration of the game

$$d_z = \left(\frac{-a}{q-p}\right) \frac{\left(\frac{q}{p}\right)^z - 1}{\left(\frac{q}{p}\right)^a - 1} + \frac{z}{q-p}$$

$$= \frac{-100}{-0.2} \left(\frac{\left(\frac{2}{3}\right)^5 - 1}{\left(\frac{2}{3}\right)^{100} - 1}\right) + \frac{5}{-(0.2)} = 409$$

★ **Markov process:-**

A stochastic process $\{x(t),\ t \geq 0\}$ is called a markov process if $p[x(t_n+1) \leq x_{n+1} / x(t_n) = x_n,\ x(t_{n-1}) = x_{n-1}$ ------

$x(t_0) = x_0]$

$= p[x(t_{n+1}) \leq x_{n+1} / x(t_n) = x_n]$

Markov chain

A stochastic process $\{x_n : n=1,2 \ldots\}$ is called markov chain, if for $j, K, J_1, \ldots J_{n-1} \in N$ (or) any subset of I_y.

$$P[x_n = K / x_{n-1} = j, x_{n-2} = j_1 \ldots x_0 = j_{n-1}]$$

$$P[x_n = K / x_{n-1} = j] = P_{jK}$$

Where $j_1, j_2 \ldots$ are called the states of the markov chain.

→ If the Transition probability P_{jK} is independent of n; The markov chain is called <u>Homogenous markov chain</u>.

→ If the Transition probability P_{jK} is dependent on n; The chain is said to be <u>Non - Homogenous markov chain</u>.

The Transition probability P_{jK} refers to the states (j, K) at two successive Trails. The transition is one - step and P_{jK} is called <u>one step (or) unit step transition probability</u>.

Multiple steps Transition probability :-

In the more general case, we are concered with the pair of states (j, K) at two non - successive trails, say, state j at the n^{th} trail and state K at the $(n+m)^{th}$ trial. The corresponding transition probability is called m - step transition probability and is denoted by $P_{jK}^{(m)}$.

$$\text{i.e } P_{jK}^{(m)} = P[x_{n+m} = K / x_n = j]$$

Probability vector :-

A probability vector is a vector (a row or column matrix) which is non-negative and all elements adding upto unity. If $p_1, p_2 \cdots p_n$ be the set of 'n' probabilities of a variable, Then the probability vector will be

$$p = [p_1\ p_2 \cdots p_n] \text{ (or) } p = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} \text{ with } \sum_{i=1}^{n} p_i = 1 \text{ and } 0 < p_i < 1$$

$\forall\ i = 1, 2 \ldots n.$

Transition matrix (or) Matrix of Transition probabilities.

The Transition probabilities p_{jk} satisfy

(i) $p_{jk} \geq 0$

(ii) $\sum_{k} p_{jk} = 1\ \forall\ j$

These probabilities may be written in the matrix form

$$p = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

This is called the tranistion probability matrix (or) matrix of transition probabilities (t.p.m) of the markov-chain.

Stochastic Matrix :-

A stochastic matrix 'p' is a square matrix with non-negative elements and unit row sums.

* If P & Q are Stochastic matrices then product PQ is also a Stochastic matrix. Thus P^n is a Stochastic matrix for all positive integer values of n. $(\forall n \in Z^+)$

Order of a Markov Chain :-

A markov chain $\{X_n\}$ is said to be of order s $(s=1,2,\ldots)$ if for all n.

$$P[X_n = k / X_{n-1} = j, X_{n-2} = j_1, \ldots X_{n-s} = j_{s-1}]$$

$$= P(X_n = k | X_{n-1} = j, \ldots X_{n-s} = j_{s-1}]$$

When ever the L.H.S is defined.

→ A markov chain $\{X_n\}$ is said to be of order - one (or Simply a markov chain) if.

$$P[X_n = k | X_{n-1} = j, X_{n-2} = j_1, \ldots)$$

$$= P[X_n = k / X_{n-1} = j]$$

$$= P_{jk}$$

Whenever $P[X_{n-1} = j, X_{n-2} = j_1 \ldots] \geq 0$

→ A chain is said to be of order zero if $P_{jk} = P_k \; \forall j$

This implies independence of X_n & X_{n-1}.

Finite markov chain :-

A markov chain $\{X_n, n \geq 0\}$ with K states, where K is finite, is said to be a finite markov chain.

The transition matrix p is, in this case, a square with K rows and K columns.

Ex :- 1) Apratice performs a random walk with absorbing barriers at 0 & 4. when ever it is at any position r $(0<r<4)$, it moves to $r+1$ with probability p or to $(r-1)$ with probability q & $p+q=1$. But as soon as it reaches 0 or 4. it remains there itself. Let X_n be the position of the partical after n moves. The different States of X_n are the different positions of the particle. $\{X_n\}$ is a markov chain whose unit-step transition probabilities are given by:

$$P[X_n = r+1 / X_{n-1} = r] = p$$

$$P[X_n = r-1 / X_{n-1} = r] = q \qquad 0 < r < 4$$

$$P[X_n = 0 / X_{n-1} = 0] = 1$$

$$P[X_n = 4 / X_{n-1} = 4] = 1$$

The Transition matrix is given by

$$\begin{array}{cc} & \text{States of } X_n \\ & \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} \text{States of} \\ X_{n-1} \end{array} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Denumerably infinite (or) Countably infinite markov chain.

The no. of states could however be infinite when the possible values of x_n form denumerable set, Then the markov chain to be denumerably infinite (or) Countably infinite and the chain is said to have a countable state space.

Ex:- Suppose that a coin with probability p for a head is tossed indefinitely. Let x_n; The out come of the n^{th} trial be k: where $k(=0,1,\dots n)$ denotes that there is a run of k successes. i.e the length of the uninterrupted block of heads is k. $\{x_n, n \geq 0\}$ constitutes a markov chain with unit step transition probabilities

$$p_{jk} = P\left[x_n = k / x_{n-1} = j\right] = \begin{cases} p, & k = j+1 \\ q, & k = 0 \\ 0, & \text{otherwise} \end{cases}$$

The Transition matrix is given by

State of X_n

States of X_{n-1}

$$\begin{array}{cc} & \begin{array}{cccccc} 0 & 1 & 2 & \cdots & k & k+1 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \\ k \\ \vdots \end{array} & \begin{bmatrix} q & p & 0 & \cdots & 0 & 0 & \cdots \\ q & 0 & p & \cdots & 0 & 0 & \cdots \\ q & 0 & 0 & \cdots & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \\ q & 0 & 0 & \cdots & 0 & p & \cdots \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots \end{bmatrix} \end{array}$$

Markov chains as Graphs

The state of markov chain may be represented by the vertices (nodes) of the graph & one-step transitions b/w states by directed arcs: if $i \to j$, then the ~~b/w~~ vertices i & j are joined by directed arc with arrow towards j. The value of P_{ij}, which corresponds to the arc weight, may be indicated in the directed arc. If $S = \{1, 2 \dots m\}$ is the set vertices corresponding to the state space of the chains 'a' is the set of directed arcs b/w these vertices;

Then the graph $G = \{S, a\}$ is directed graph (or) digraph (or) Transition graph of the chain. A digraph such that its arc weights are positive & sum of the arc weights of the arcs from each node unity is called a Stochastic graph; The diagraph of a markov chain is a stochastic graph.

— A path in a diagraph is any sequence of arcs where the final vertex of one arc is the initial vertex of the next arc.

Higher Transition probabilities.

* **Chapman - kolmogorov theorem.**

If p is the tpm of a homogeneous markov chain, Then the n- step tpm $p(n)$ is equal to p^n

$$\text{i.e } \left[p_{ij}^{(n)}\right] = \left[p_{ij}\right]^n.$$

* **Classification of states & chains**

The states j, $j = 0, 1, 2 \ldots$ of a markov chain $\{x_n, n \geq 0\}$ can be classified in a distinctive manner According to some fundamental properties of the system.

Communication relation

1) If $p_{ij}^{(n)} > 0$ for some $n \geq 1$, then we say that state j is accessible from state i, the relation is denoted

by $i \to j$ conversely, if $\forall n$; $p_{ij}^{(n)} = 0$, Then j is not accessible from i, in notation $i \nrightarrow j$

2) Two accessible States i & j are said to be communicate State i communicates with itself $\forall i \geq 0$. If the state i communicates with state j & state j communicates with state k; Then state i communicates with state k. Two States that communicate are in The same class. A state is called an essential state. If it communicates with every state it leads so.

3) If $p_{ij}^{(n)} > 0$ for some n & $\forall i$ & j. Then every state can be reached from every other state. Then the markov chain is said to be irreducible. Then the transition matrix is irreducible. otherwise the chain is said to be reducable (or) non-irreducible.

4) A State i is said to be an absorbing state. if and only if $p_{ii} = 1$. A markov chain is absorbing if it has at least one absorbing state & it is possible to go from every non-absorbing state to atleast one absorbing state in one (or) more steps.

5) periodicity :- A state i is a returnstate if $P_{ij}^{(n)} > 0$ for some $n \geq 1$. The period d_i of a return to state i is defined as the greatest common divisor of all m such that $P_{ij}^{(m)} > 0$. Thus

$$d_i = G.C.D \{m : P_{ij}^{(m)} > 0\}$$

A state i is said to aperiodic if $d_i = 1$ & periodic if $d_i > 1$. Clearly state i is aperiodic if $P_{ij} \neq 0$.

Classification of States

7) The probability that the chain starting from state i returns to states i for the first time at the n^{th} step (or) after n (transitions) is denoted by $f_{ii}^{(n)}$, $n = 1, 2 \ldots$ and is called as First time return time probability (or) the recurrence time probability.

If $F_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)} = 1$, the return to state i is certain & $M_{ii} = \sum_{n=1}^{\infty} n f_{ii}^{(n)}$ is called the mean recurrence time of the state i.

Persistent (or) recurrent state :-

A state i is said to be persistent (or) recurrent i if $F_{ii} = 1$

Transient State :-

The state "i" is said to be transient if $F_{ii} < 1$

(i.e the return to state i is uncertain)

Null persistent & Non-Null persistent State :-

The state "i" is said to be non-null persistent (or) positive persistent if μ_{ii} is finite. The state "i" is said to be Null persistent if $\mu_{ii} = \infty$.

Classification of Chains

8) Ergodic :-

Apolstic recurrent and aperiodic state of a Markov Chain is called Ergodic. A markov chain all of whose states are ergodic said to be a ergodic chain.

9) A tPM, P is said to be a regular matrix if all the entries of power P^m $(m = 2, 3 \ldots)$ are positive. A homogenous markov chain is said to be regular chain if its tPM is a regular matrix.

A tPM, P is said to be stochastic matrix, if the Element of each of the rows are non-negative and the sum of elements in each row is equal to 1.

10) A homogeneous markov chain will have a tpm. That is independent of initial state i & steps n as $n \to \infty$ and is called steady state probability.

$$q_j = \lim_{n \to \infty} p_{ij}^{(n)} \quad \& \quad \Sigma q_j = 1 \ \forall j$$

where q_j is called limiting state probability and is interpreted as the long-run proportion of time the markov chain spends in state j.

Theorem :- ★

A Stochastic matrix p is not regular if 1 Occures in the principle main diagonal.

1. which of the following matrices are stochastic.

(i) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (ii) $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ (iii) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ (iv) $\begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}$ (v) $\begin{bmatrix} 0 & 2 \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}$

Sol (i) The matrix is a square matrix with non-negative entries and sum of elements in each row is equal to 1

$\therefore \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ & $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ are stochastic matrices

(iii) It is not a square matrix

$\therefore$ It is not stochastic.

(iv) The matrix is not stochastic, because it contains negative elements

(v) The matrix is square matrix but sum in each row is not equal to 1

$\therefore$ It is not a stochastic matrix.

8) Which of the Stochastic matrices are regular.

(i) $A = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$ (ii) $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$ (iii) $\begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$

(iv) $\begin{bmatrix} \frac{1}{2} & \frac{1}{4} & 1 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 1 \end{bmatrix}$

Sol (i) G.T $A = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$

Not Regular since 1 lies on the main diagonal

(ii) G.T $B = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$

$B^2 = B \cdot B = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 3/8 & 3/8 & 0 \end{bmatrix}$

$B^3 = B^2 \cdot B = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{7}{16} & \frac{7}{16} & \frac{1}{8} \end{bmatrix}$

Since entries b_{13}, b_{23} are zero

$\therefore$ B is not Regular.

(iii) G.T $C = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$

$\therefore C^2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$, $C^3 = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$

$C^4 = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$, $C^5 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{8} & \frac{1}{2} & \frac{3}{8} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$

Since all entries of some powers of C are positive

$\therefore$ C is regular stochastic process.

(iv) G.T

$D = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & 1 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 1 \end{bmatrix}$

Not Regular since 1 lies on the main diagonal.

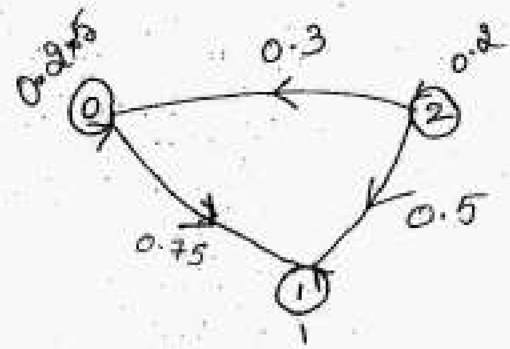
3) Represent the following transition matrices as a diagraph

(i) $\begin{bmatrix} 0.25 & 0.75 & 0 \\ 0 & 1 & 0 \\ 0.3 & 0.5 & 0.2 \end{bmatrix}$ (ii) $\begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{3}{4} & \frac{1}{4} \end{bmatrix}$

Sol:- (i) G.T the tPM of markov chain is

$$\begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & 0.25 & 0.75 & 0 \\ 1 & 0 & 1 & 0 \\ 2 & 0.3 & 0.5 & 0.2 \end{array}$$

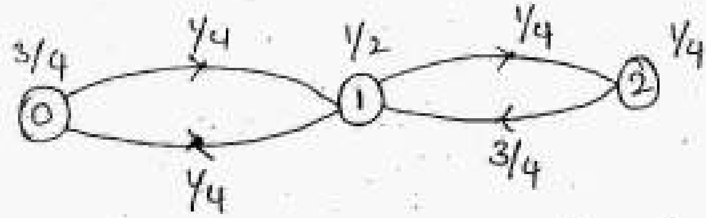
The diagraph of chain is



(ii) G.T, TPM of markov chain is

$$\begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{3}{4} & \frac{1}{4} \end{bmatrix}$$

∴ The diagraph of chain is



4) Suppose that the probability of a dry day (state 0) following a rainy day (state 1) is $\frac{1}{3}$ and that the probability of a rainy day following a dry day is $\frac{1}{2}$. Given that may 1^{st} is a dry day. Find the probability (i) may 3^{rd} is also a dry day & (ii) may 5 is also a dry day.

dry day = 0
rainy day = 1

Sol:- we have two - state markov chain such that

$P_{10} = \frac{1}{3}$ & $P_{01} = \frac{1}{2}$ & t.P.M is given by

$$P = \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & \frac{1}{3} & \frac{2}{3} \end{array}$$

we have $p^2 = \begin{bmatrix} \frac{5}{12} & \frac{7}{12} \\ \frac{7}{18} & \frac{11}{18} \end{bmatrix}$

$$p^4 = \begin{bmatrix} \frac{173}{432} & \frac{259}{432} \\ \frac{259}{648} & \frac{389}{648} \end{bmatrix}$$

Given that, may 1 is a ~~dry~~ day, then the probability that may 3 is a dry day is $\frac{5}{12}$ & that may 5 is a dry day is $\frac{173}{432}$

*Imp

5) A training process is considered as a two-state markov chain. If it rains, it is considered to be in state 0 & it does not rain, The chain is in the state of 1. The transition probability of the markov chain is defined by $p = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$. Find the probability that it will rain for 3 days from today assuming that it is raining today. Assume that the initial probabilities of state 0 (or) state 1 as 0.4 & 0.6 respectively.

<u>Sol</u>:- G.T the one step t.p.m is given by

$$P = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$$

$$P^2 = P \cdot P$$

$$= \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0.44 & 0.56 \\ 0.28 & 0.72 \end{bmatrix}$$

$$P^3 = P^2 \cdot P = \begin{bmatrix} 0.44 & 0.56 \\ 0.28 & 0.72 \end{bmatrix} \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 0.376 & 0.624 \\ 0.312 & 0.688 \end{bmatrix}$$

The probability that is will rain on 3^{rd} day, given that it will rain today is 0.376.

6) Let $\{X_n, n \geq 0\}$ be a markov chain with Three states $\{0,1,2\}$ & with transition matrix

$$\begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & \frac{3}{4} & \frac{1}{4} & 0 \\ 1 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 2 & 0 & \frac{3}{4} & \frac{1}{4} \end{array}$$

& the initial distribution $P\{X_0 = i\} = \frac{1}{3}$; $i = 0,1,2$. Find $P\{\overset{F}{X_3}=1, \overset{present}{X_2=2}, \underbrace{X_1=1, X_0=2}_{past}\}$

Sol:- G.T, the tpm of the Markov Chain is

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{3}{4} & \frac{1}{4} \end{bmatrix} \end{array} \quad \& \quad P\{X_0 = i\} = \frac{1}{3}, \; i = 0,1,2$$

Then we have

$$P\{X_1 = 1 / X_0 = 2\} = \frac{3}{4} = P_{21}$$

$$P\{X_2 = 2 / X_1 = 1\} = \frac{1}{4} = P_{12}$$

$$P\{x_2=2,\ x_1=1/x_0=2\} = P\{x_2=2/x=1\}\cdot P\{x_1=1/x_0=2\}$$

$$= \frac{1}{4}\cdot\frac{3}{4} = \frac{3}{16}$$

$$\therefore P\{x_2=2,\ x_1=1,\ x_0=2\} = P\{x_2=2,\ x_1=1/x_0=2\}\cdot P\{x_0=2\}$$

$$= \cancel{\frac{3}{16}}\cdot\frac{1}{\cancel{3}} = \frac{1}{16}$$

$$\therefore P\{x_3=1,\ x_2=2,\ x_1=1,\ x_0=2\} = P\{x_3=1/x_2=2,\ x_1=1,\ x_0=2\}$$

$$P\{x_2=2,\ x_1=1,\ x_0=2\}$$

$$= P\{x_3=1/x_2=2\}\cdot P\{x_2=2,\ x_1=1,\ x_0=2\}$$

$$= \frac{3}{4}\cdot\frac{1}{16} = \frac{3}{64}$$

*) The t.p.m of a markov chain $\{x_n\}$, $n=1,2,3\ldots$ having 3 states 1, 2 & 3 is $P = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$ and the initial distribution is $P^{(0)} = (0.7, 0.2, 0.1)$ find (i) $P\{x_2=3\}$ (ii) $P\{x_3=2,\ x_2=3,\ x_1=3,\ x_0=2\}$.

Sol G.T the t.pm of markov chain is

$$P = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} \end{array}$$

We have $P(x_0=1)=0.7$, $P(x_0=2)=0.2$, $P(x_0=3)=0.1$

Also $p^{(2)} = P^2 = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$

$$P^2 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} \quad 1 & \quad\quad 2 & \quad\quad 3 \end{array} \\ \begin{bmatrix} 0.43 & 0.31 & 0.26 \\ 0.24 & 0.42 & 0.34 \\ 0.36 & 0.35 & 0.29 \end{bmatrix} \end{array}$$

(i) $P(X_2 = 3) = \sum_{i=1}^{3} P\{X_2 = 3 / X_0 = i\} \cdot P\{X_0 = i\}$

$= P\{X_2 = 3 / X_0 = 1\} \cdot P\{X_0 = 1\} + P\{X_2 = 3 / X_0 = 2\}$

$P\{X_0 = 2\} + P\{X_2 = 3 / X_0 = 3\} \cdot P\{X_0 = 3\}$

$= P_{13}^{(2)} \cdot P\{X_0 = 1\} + P_{23}^{(2)} P\{X_0 = 2\} + P_{33}^{(2)} \cdot P\{X_0 = 3\}$

$= 0.26\,(0.7) + 0.34\,(0.2) + 0.29\,(0.1)$

$= 0.279.$

(ii) $P\{X_1 = 3 / X_0 = 2\} = P_{23} = 0.2$

$P\{X_1 = 3,\ X_0 = 2\} = P\{X_1 = 3 / X_0 = 2\} \cdot P\{X_0 = 2\}$

$= 0.2 \times 0.2$

$= 0.04$

$P\{X_2 = 3,\ X_1 = 3,\ X_0 = 2\} = P\{X_2 = 3 / X_1 = 3,\ X_0 = 2\} \cdot P\{X_1 = 3, X_0 = 2\}$

$= P\{X_2 = 3 / X_1 = 3\} \cdot P\{X_1 = 3,\ X_0 = 2\}$

$= 0.3 \times 0.04$ (by markov property)

$= 0.012$

$P\{X_3=2, X_2=3, X_1=3, X_0=2\}$

$= P\{X_3=2 / X_2=3, X_1=3, X_0=2\} \cdot P\{X_2=3, X_1=3, X_0=2\}$

$= P\{X_3=2 / X_2=3\} \cdot P\{X_2=3, X_1=3, X_0=2\}$

$= 0.4 \times 0.012$

$= 0.0048$

Imp

8) A fair die tossed repeatedly. If x_n denotes the maximum of the numbers occuring in the first n tosses. Find the transition probability matrix P of the markov chain $\{x_n\}$. Find also P^2 and $P(X_2=6)$

Sol:- State Space $= \{1,2,3,4,5,6\}$

The tpm is formed using the following analysis

Let X_n = The maximum of the numbers occurring in the First n trials $= 3$ (say)

Then $X_{n+1} = 3$ if the $(n+1)^{th}$ trail results is 1, 2 or 3.

$= 4$ if the $(n+1)^{th}$ trail results is 4

$= 5$ if the $(n+1)^{th}$ trail results is 5

$= 6$ if the $(n+1)^{th}$ trail results is 6

$\therefore P\{X_{n+1}=3 / X_n=3\} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6}$

$\therefore P\{X_{n+1}=i / X_n=3\} = \frac{1}{6}$ when $i = 4, 5, 6$

$\therefore$ The tPm of chain is

x_{n+1}

$$
P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}
\begin{array}{c}
\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\begin{bmatrix}
\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\
0 & \frac{2}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\
0 & 0 & \frac{3}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\
0 & 0 & 0 & \frac{4}{6} & \frac{1}{6} & \frac{1}{6} \\
0 & 0 & 0 & 0 & \frac{5}{6} & \frac{1}{6} \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{array}
$$

x_n

$\frac{6}{6} = 1$

$p^{(0)}$ $p^{(1)}$ $p^{(2)}$ (or) p^2

(ii) $$P^2 = \frac{1}{36}\begin{bmatrix}
1 & 3 & 5 & 7 & 9 & 11 \\
0 & 4 & 5 & 7 & 9 & 11 \\
0 & 0 & 9 & 7 & 9 & 11 \\
0 & 0 & 0 & 16 & 9 & 11 \\
0 & 0 & 0 & 0 & 25 & 11 \\
0 & 0 & 0 & 0 & 0 & 36
\end{bmatrix}$$

Initial state probability distribution is $p^{(0)} = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$

(Since all the values of 1, 2, ... 6 are equally likely)

(iii) $$p(X_2 = 6) = \sum_{i=1}^{6} p(X_2 = 6 / X_0 = i)\; p(X_0 = i)$$

$$= \frac{1}{6}\sum_{i=1}^{6} P_{i6}^{(2)}$$

$$= \frac{1}{6}\left[P_{16}^{(2)} + P_{26}^{(2)} + P_{36}^{(2)} + P_{46}^{(2)} + P_{56}^{(2)} + P_{66}^{(2)}\right]$$

$$= \frac{1}{6}\left[\frac{11}{36} + \frac{11}{36} + \frac{11}{36} + \frac{11}{36} + \frac{11}{36}\right]$$

$$= \frac{1}{6\times 36}\left[11 + 11 + 11 + 11 + 11 + 36\right] = 91$$

$$P(X_2=6) = \frac{91}{216}$$

9) A gambler has Rs. 2 He bets Rs. 1 at a time and wins Rs. 1 with probability $\frac{1}{2}$. He stops playing if he loses Rs. 2 (or) wins Rs. 4.

(a) What is the tpm of related markov chain?

(b) What is the probability that he has lost has money at the end of 5 plays?

(c) What is the probability that the game lasts more than 7 plays?

Sol Let X_n represent the amount with the player at the end of the n^{th} round of the play.

The state space $X_n = \{0,1,2,3,4,5,6\}$

When the game is stopped, if the player loses Rs. 2, $X_n = 0$ (or) If he wins Rs. 4. $X_n = 6$.

The tpm of markov chain is written as

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

This markov chain is called random walk with absorbing barriers at 0 & 6, since the chain cannot come out of the states 0 & 6, once it has entred.

b) Since the player has Rs. 2 to start the play, the initial probability distribution of x_n is

$$p^{(0)} = (0\ 0\ 1\ 0\ 0\ 0)$$

probability distribution after one play is given by

$$p^{(1)} = p^{(0)} \cdot p = (0\ 0\ 1\ 0\ 0\ 0\ 0)p = (0\ \tfrac{1}{2}\ 0\ \tfrac{1}{2}\ 0\ 0\ 0)$$

$$p^{(2)} = p^{(1)} \cdot p = (0\ \tfrac{1}{2}\ 0\ \tfrac{1}{2}\ 0\ 0\ 0)p = (\tfrac{1}{4}\ 0\ \tfrac{1}{2}\ 0\ \tfrac{1}{4}\ 0\ 0)$$

$$p^{(3)} = p^{(2)} \cdot p = (\tfrac{1}{4}\ 0\ \tfrac{1}{2}\ 0\ \tfrac{1}{4}\ 0\ 0)p = (\tfrac{1}{4}\ \tfrac{1}{4}\ 0\ \tfrac{3}{8}\ 0\ \tfrac{1}{8}\ 0)$$

$$p^{(4)} = p^{(3)} \cdot p = (\tfrac{1}{4}\ \tfrac{1}{4}\ 0\ \tfrac{3}{8}\ 0\ \tfrac{1}{8}\ 0)p = (\tfrac{3}{8}\ 0\ \tfrac{5}{16}\ 0\ \tfrac{1}{4}\ 0\ \tfrac{1}{16})$$

$$p^{(5)} = p^{(4)} \cdot p = (\tfrac{3}{8}\ 0\ \tfrac{5}{16}\ 0\ \tfrac{1}{4}\ 0\ \tfrac{1}{16})p = (\tfrac{3}{8}\ \tfrac{5}{32}\ 0\ \tfrac{9}{32}\ 0\ \tfrac{1}{8}\ \tfrac{1}{6})$$

The probability that the player has lost his money at the end of 5 plays $= p(X_5 = 0) = \frac{3}{8}$

c) The probability that the game lasts more than 7 plays

= p (system is neither in state 0 nor in 6 at the end of the seventh round).

using $p^{(6)} = p^{(5)} \cdot p = \left(\frac{29}{64} \quad 0 \quad \frac{7}{32} \quad 0 \quad \frac{13}{64} \quad 0 \quad \frac{1}{8}\right)$

$$\begin{array}{cc} & \begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ p^{(7)} = p^{(6)} \cdot p = & \left(\begin{array}{ccccccc} \frac{29}{64} & \frac{7}{64} & 0 & \frac{27}{128} & 0 & \frac{13}{128} & \frac{1}{8} \end{array}\right) \end{array}$$

$$\therefore P\{x_7 = 1, 2, 3, 4 \text{ or } 5\} = \frac{7}{64} + 0 + \frac{27}{128} + 0 + \frac{13}{128}$$

$$= \frac{27}{64}$$

10) Three boys A, B & C are throwing a ball to each other. A always throws the ball to B & B always throws the ball to C; but C is just as likely to throw the ball to B as A. Show that the process is markovian. Find the transition matrix & classify the states. Do all the states are ergodic

Sol:- The transition probability matrix of the process $\{x_n\}$ is given below.

States of x_n (future)

States of x_{n-1}

$$P_{(\text{present})} = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \end{array}$$

States of x_n depends only on states of x_{n-1} but not on states of $x_{n-2}, x_{n-3}, \ldots$ or earlier states.

$\therefore \{x_n\}$ is a markov chain.

Now $P^2 = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$, $P^3 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$

$P_{11}^{(3)} > 0$, $P_{13}^{(2)} > 0$, $P_{21}^{(2)} > 0$, $P_{22}^{(2)} > 0$, $P_{33}^{(2)} > 0$ and all other

∴ The chain is irreducible. Irreducible ;—

$P^4 = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$ $P^5 = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{8} & \frac{3}{8} & \frac{1}{2} \end{bmatrix}$ a/constable $P_{ij}^{(n)} > 0$.

$P^6 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{8} & \frac{1}{2} \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} \end{bmatrix}$ & so on.

We note that $P_{ii}^{(2)}, P_{ii}^{(3)}, P_{ii}^{(4)}, P_{ii}^{(5)}, P_{ii}^{(6)}$ etc are >0. for $i=2,3$

G.C.D of $2, 3, 4, 5, 6 \ldots = 1$

The states 2 & 3 (i.e B & C) are periodic with period 1

i.e aperiodic.

We note that $P_{ii}^{(3)}, P_{ii}^{(5)}, P_{ii}^{(6)}$ etc, are >0 & G.C.D of

$3, 5, 6 \ldots = 1$

∴ The state 1 (i.e state A) is periodic with period 1 i.e

aperiodic.

Since the chain is finite and irreducible, all its

States are non-null persistent.

Hence, all the states are ergodic.

11) Find the nature of states of the markov chain with tPM

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} \end{array}$$

Sol:- G.T the time of markov chain is

$$P = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$$

$$P^2 = P \cdot P = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

$$P^3 = P^2 \cdot P = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} = P$$

$$P^4 = P^3 = P = P \cdot P = P^2$$

$$\text{lly } P^5 = P^3 \cdot P^2 = P \cdot P^2 = P$$

$$P^6 = P^4 \cdot P^2 = P^2 \cdot P^2 = P^2$$

In general $P^{2n} = P^2$ & $P^{2n+1} = P^{2n} \cdot P = P^2 \cdot P = P^3 = P$

Also, we observe

$P_{00}^{(2)} > 0$, $P_{02}^{(2)} > 0$, $P_{11}^{(2)} > 0$, $P_{20}^{(2)} > 0$, $P_{22}^{(2)} > 0$.

$P_{01}^{(1)} > 0$, $P_{10}^{(1)} > 0$, $P_{12}^{(1)} > 0$, $P_{21}^{(1)} > 0$.

Hence the markov chain is irreducible.

Also $P_{ii}^{(2)} > 0$, $P_{ii}^{(4)} > 0$, $P_{ii}^{(6)} \geq 0$ & so on for every 'i'.

Thus $P_{ii}^{(2n)} > 0$, $P_{ii}^{(2n+1)} = 0$, for each i.

$\therefore$ All the states of the markov chain are periodic with period 2.

Since the markov chain is finite & irreducible, all its states are non-null persistent. All the states are not ergodic.

* <u>Irreduable</u>:- If $P_{ij}^{(n)} > 0$ for some n & $\forall$ i & j. Then every state can be reached from every other state. When this condition is satisfied. The markov chain is said to be irreducible. The tpm of an irreducible chain is an irreducible-matrix otherwise the chain is said to be reduciable.

12) The transition probability matrix of a markov chain is given by $\begin{bmatrix} 0.3 & 0.7 & 0 \\ 0.1 & 0.4 & 0.5 \\ 0 & 0.2 & 0.8 \end{bmatrix}$. Is this matrix irreducible.

Sol Consider the Three States as 0, 1, 2

$$\begin{array}{c c} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \end{array} & \begin{bmatrix} 0.3 & 0.7 & 0 \\ 0.1 & 0.4 & 0.5 \\ 0 & 0.2 & 0.8 \end{bmatrix} \end{array}$$

In this chain we go from state 0 to state 1 with probability of 0.7 & from state 1 to state 2 with probability 0.5. Thus it is possible to go from state 0 to state 2.

∴ The chain is irreducible & all the states are recurrent.

3 x. If the matrix $\begin{bmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 \\ 0.2 & 0.4 & 0.1 & 0.3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ irreducible.

Sol
$$P = \begin{bmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 \\ 0.2 & 0.4 & 0.1 & 0.3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$P^2 = P \cdot P = \begin{bmatrix} 0.34 & 0.66 & 0 & 0 \\ 0.33 & 0.67 & 0 & 0 \\ 0.22 & 0.44 & 0.01 & 0.33 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 0.334 & 0.666 & 0 & 0 \\ 0.333 & 0.667 & 0 & 0 \\ 0.222 & 0.444 & 0.001 & 0.003 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

W.K.T, one of the rule to prove that, a matrix is irreducible is that the sum of each row of the matrix must be equal to 1, but in matrix P^3, the sum of third row is not equal to 1. Hence The given matrix is not irreducible.

* If $X = [x_1, x_2 \dots x_n]$ is the Steady-State distribution of the chain whose tpm is n^{th} Order Square matrix 'p'.

Then $X = XP$.

14) If the tpm of a markov chain is $\begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$. Find the steady State distribution of the chain.

Sol Let $X = [x_1 \ x_2]$ is the steady-state distribution of the chain

Then $X = XP$ and $x_1 + x_2 = 1$ — ①

$$[x_1 \ x_2] = [x_1 \ x_2] \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$[x_1 \ x_2] = \left[\tfrac{1}{2}x_2 \quad x_1 + \tfrac{1}{2}x_2\right]$$

$$x_1 = \tfrac{1}{2}x_2 \Rightarrow 2x_1 = x_2 \text{ — ②}$$

$$x_2 = x_1 + \tfrac{1}{2}x_2 \text{ — ③}$$

sub eq^n ② in ①, we get

$$x_1 + 2x_1 = 1$$

$$3x_1 = 1$$

$$x_1 = \frac{1}{3}$$

from eq^n ②, we get

$$x_2 = 2 \times \frac{1}{3}$$

$$= \frac{2}{3}$$

$\therefore$ Hence $\begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}$ be the steady state distribution of the chain.

15) Study of passage of english text to find a vowel followed by a vowel or a consonant followed by a consonant or a vowel reveal the following transition probability matrix $\begin{bmatrix} 0.12 & 0.88 \\ 0.54 & 0.46 \end{bmatrix}$. Find the percentage of letters in the text book. which are expected to be vowels.

Sol Given that the transition probability matrix $p = \begin{bmatrix} 0.12 & 0.88 \\ 0.54 & 0.46 \end{bmatrix}$

we have to find the limiting probabilities

$$X = XP \ \& \ x_1 + x_2 = 1 \quad \text{---} \ (1)$$

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 0.12 & 0.88 \\ 0.54 & 0.46 \end{bmatrix}$$

$$x_1 = 0.12x_1 + 0.54x_2 \quad \text{---} \ (2)$$

$$x_2 = 0.88x_1 + 0.46x_2 \quad \text{---} \ (3)$$

Sub, $x_2 = 1 - x_1$ in (2) (from (1))

we, get

$$0.12x_1 + 0.54(1 - x_1) = x_1$$

$$\Rightarrow 0.12x_1 - 0.54x_1 + 0.54 = x_1$$

$$\Rightarrow x_1 = 0.3803$$

$$\therefore x_2 = 1 - x_1$$

$$= 1 - 0.3803$$

$$x_2 = 0.6197$$

$\therefore$ Hence, the percentage of letters in the textbook which are expected to be vowels is 38%.

16) The School of International studies for population found out by its survey that the mobility of the population of a state to village, town and city is the following percentage.

From \ To	Village	Town	City
Village	30%	20%	50%
Town	30%	50%	20%
City	10%	40%	50%

What will the proportion of population in village, town & city after two years? present population has proportion of 0.4, 0.3 and 0.3 village, town and city respectively. Find the proportions in the long run.

Sol:- Given that the school of International studies for population found to its survey that the mobility of the population of a state to village, Town & City are.

From \ To	village	Town	city
village	30%	20%	50%
Town	30%	50%	20%
city	10%	40%	50%

Which can be written as

$$P = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

The present population proportion $P_0 = [0.4, 0.3, 0.3]$

The proportion of population in village, town & city after one year is

$$P_1 = P_0 \cdot P$$

$$= \begin{bmatrix} 0.4 & 0.3 & 0.3 \end{bmatrix} \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

$$P_1 = \begin{bmatrix} 0.24 & 0.35 & 0.4 \end{bmatrix}$$

Hence the proportion of population in village, town and city after one year

$$P_1 = \begin{bmatrix} 0.24 & 0.35 & 0.4 \end{bmatrix}$$

→ The proportion of population in village, town and city after two years

$$P_2 = P_1 \cdot P$$

$$= \begin{bmatrix} 0.24 & 0.35 & 0.4 \end{bmatrix} \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.218 & 0.387 & 0.395 \end{bmatrix}$$

Hence the proportion of population in village, town and city after two years are 0.218, 0.387 & 0.395.

(b) The proportion of population in village, town and city in long run.

i.e the steady-state distribution of the chain, then

$$X = XP \text{ and } x_1 + x_2 + x_3 = 1$$

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

$$x_1 = 0.3x_1 + 0.3x_2 + 0.1x_3$$

$$\Rightarrow 0.7x_1 = 0.3x_2 + 0.1x_3 \quad \text{— ①}$$

$$x_2 = 0.2x_1 + 0.5x_2 + 0.4x_3$$

$$\Rightarrow 0.5x_2 = 0.2x_1 + 0.4x_3 \quad \text{— ②}$$

$$x_3 = 0.5x_1 + 0.2x_2 + 0.5x_3$$

$$\Rightarrow 0.5x_3 = 0.5x_1 + 0.2x_2 \quad \text{— ③}$$

$$x_1 + x_2 + x_3 = 1$$

$$x_1 = 1 - x_2 - x_3 \quad \text{— ④}$$

Now multipyling eq^n ④ with 0.7 and subtracting eq^n ① we get,

$$0.7x_1 = 0.7 - 0.7x_2 - 0.7x_3$$

$$\underline{0.7x_1 = 0 + 0.3x_2 + 0.1x_3}$$

$$0 = 0.7 - 1.0x_2 - 0.8x_3$$

$$\therefore x_2 + 0.8x_3 = 0.7 \quad \text{— ⑤}$$

Now multiply eq^n ② with 0.5 and eq^n ③ with 0.2 then we get

$$0.25x_2 = \cancel{0.1x_1} + 0.2x_3$$

$$\underline{0.1x_3 = \cancel{0.1x_1} + 0.04x_2}$$

$$0.25x_2 - 0.1x_3 = 0.2x_3 - 0.04x_2$$

$$0.25x_2 + 0.04x_2 = 0.2x_3 + 0.1x_3$$

$$0.29x_2 = 0.3x_3$$

$$x_2 = \left(\frac{0.3}{0.29}\right)x_3$$

$$x_2 = (1.034)x_3 \quad \text{—— ⑥}$$

Now, substituting eq^n ⑥ in eq^n ⑤; we get

$$(1.034)x_3 + 0.8x_3 = 0.7$$

$$1.834x_3 = 0.7$$

$$x_3 = \frac{0.7}{1.834}$$

$$x_3 = 0.382$$

Sub $x_3 = 0.382$ in eq^n ⑥, we get

$$x_2 = (1.034)(0.382)$$

$$x_2 = 0.394$$

put $x_2 = 0.394$ and $x_3 = 0.382$ in eq^n ④, we get

$$x_1 = 1 - 0.394 - 0.382$$

$$x_1 = 0.224$$

$$X = [x_1 \ \ x_2 \ \ x_3] = [0.224 \quad 0.394 \quad 0.382]$$

∴ The long run proportion of population move from state to village, town and city is 22.4%, 39.4% and 38.2% respectively.

17) If the transition probability matrix of market shares of three brands A, B and C is $\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.8 & 0.1 & 0.1 \\ 0.35 & 0.25 & 0.4 \end{bmatrix}$ and the initial market shares are 50%, 25% and 25%.

Find (i) The market shares in second and third periods.

(ii) The limiting probabilities.

Sol:- Given that, the initial market shares of three brands A, B and C are 50%, 25% and 25%. which can be written as $P_0 = \begin{bmatrix} 0.5 & 0.25 & 0.25 \end{bmatrix}$

The transition probability matrix

$$P = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.8 & 0.1 & 0.1 \\ 0.35 & 0.25 & 0.4 \end{bmatrix}$$

The market shares in the second period $= P_0 \cdot P^2$

$$= \begin{bmatrix} 0.5 & 0.25 & 0.25 \end{bmatrix} \begin{bmatrix} 0.5050 & 0.225 & 0.27 \\ 0.435 & 0.275 & 0.29 \\ 0.48 & 0.23 & 0.29 \end{bmatrix}$$

$$= \begin{bmatrix} 0.4813 & 0.2387 & 0.2800 \end{bmatrix}$$

Hence the market shares in the third period $= P_0 \cdot P^3$

$$= \begin{bmatrix} 0.5 & 0.25 & 0.25 \end{bmatrix} \begin{bmatrix} 0.4145 & 0.2085 & 0.19 \\ 0.4675 & 0.2105 & 0.1985 \\ 0.4652 & 0.2307 & 0.221 \end{bmatrix}$$

$$= \begin{bmatrix} 0.4815 & 0.2382 & 0.2802 \end{bmatrix}$$

Hence the market shares in the third period is

48.1%, 23.8%, 28.02%.

ii) The limiting probability can be interpreted as the long run market shares that can be reached to an equilibrium.

$X = [x_1 \ x_2 \ x_3]$ then $X = XP$ and $x_1 + x_2 + x_3 = 1$ —— (1)

$$[x_1 \ x_2 \ x_3] = [x_1 \ x_2 \ x_3] \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.8 & 0.1 & 0.1 \\ 0.35 & 0.25 & 0.4 \end{bmatrix}$$

Now $x_1 = 0.4x_1 + 0.8x_2 + 0.35x_3$

$0.6x_1 - 0.8x_2 - 0.35x_3 = 0$ —— (2)

Now $x_2 = 0.3x_1 + 0.1x_2 + 0.25x_3$

$-0.3x_1 + 0.9x_2 - 0.25x_3 = 0$ —— (3)

Now $x_3 = 0.3x_1 + 0.1x_2 + 0.4x_3$

Solving eq^{ns} ①, ② and ③, we get

$$x = 0.4813, \quad y = 0.2383, \quad z = 0.2804$$