08/09/21
Wednesday

Syllabus

UNIT-I

Probability and Random Variables

UNIT-II

Probability Distributions

UNIT-III

Estimation of Testing of Hypothesis
(Large Samples)

UNIT-IV

Testing of Hypothesis (Small Samples)

UNIT-V

Stochastic Process and Markov Chains

UNIT-I: Probability & Random Variables

Experiment: An experiment is a physical action or process i.e observed and the result is noted.

Eg: Tossing a coin, Throwing a die

* Experiments are of two types:

Deterministic Experiment → An experiment is called deterministic experiment if the result can be predicted with certainity prior to the performance.

Eg: Throwing a stone will always touches the ground

Undeterministic Experiment → Its defined as if the result can't be predicted prior to the performance with certainity, but all the outcomes can be predicted before in-hand.

Eg: Tossing a coin The possible outcomes are Head or Tail but we can't predict the exact result.

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

Outcome or Event: Result of an experiment is called an outcome or event.

* Events are of two types:

Elementary Events → The elementary event is an event which can't be broken further.

Eg: Getting a number 5 on die

Compound Events → The compound event is an event which can be broken further into smaller events.

Eg: Getting an odd number when throwing a die

Sample Space (S): The set of all possible outcomes of a random experiment is called a Sample Space. Each point in a Sample space is called a Sample Point.

Eg: Tossing a coin

Sample Space $(S) = \{H, T\}$

Finite Sample Space → If the number of sample Points in a sample space are finite, then the sample space is called finite sample space.

Infinite Sample Space → If the number of sample Points in a sample space are infinite, then the sample space is called infinite sample space.

Equally Likely Events: A set of events are said to be equally likely, if no one of them is expected to occur in preference to the other in any single trial of an experiment.

Eg: When tossing a coin getting head or tail are equally likely events.

Mutually Exclusive Events: Events which rejects the happening of other events.

Favorable Events: The events which are favorable to a particular event of an experiment are known as Favorable Events.

Eg: Getting 2,4,6 are favorable to the event of getting an even number.

Probability: If an experiment is performed, "n" is the no. of total cases and "m" is the no. of favorable cases of an event 'A'. Then, the probability of an event 'A' is defined by

$$P(A) = \frac{\text{No. of favorable cases}}{\text{Total no. of cases}}$$

$$\boxed{P(A) = \frac{m}{n}}$$

① Find the probability of getting a head in tossing a coin.

Sol: No. of ways $= 2^n$ $n \rightarrow$ no. of coins tossed

Events $= 2$

Total events $(n) = 2 = \{H, T\}$

Probability of getting head $(m) = 1$

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

② Find the probability of getting 1 head in tossing 2 coins.

Sol: No. of events $= 2^n$

$= 2^2$

$n = 4$

$m = \{HH, HT, TH, TT\}$

$m = 2$

$P(A) = \frac{m}{n} = \frac{2}{4}$

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

③ If 3 coins are tossed. Find the probability of getting
i) Three heads
ii) No head
iii) Two heads

Sol: $n = 2^3$

$n = 8$ events

which are {HHH, HTH, THH, HHT, TTT, TTH, THT, HTT}

i) $P(A) = \frac{1}{8}$

ii) $P(A) = \frac{1}{8}$

iii) $P(A) = \frac{3}{8}$

④ Find the probability of getting a sum 9 if 2 dice are thrown.

Sol: Events $= 6^n$

$= 6^2$

$n = 36$

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

m = Sum of getting 9

= {(4,5) (5,4) (3,6) (6,3)}

$m = 4$

$P(A) = \frac{4}{36} = \frac{1}{9}$

09/09/21

⑤ Find the probability of getting 1 Red King if we select a card from a pack of 52 cards.

Sol:

52 cards

- Diamonds (13)
- Hearts (13)
- Spades (13)
- Clubs (13)

Diamonds + Hearts → Red (26) (K, Q, J, A, 1, 2 ..... 10)

Spades + Clubs → Black (26) (K, Q, J, A, 1, 2 ..... 10)

Total no. of cards $(n) = 52$

m = Probability i.e Favorable cases of getting 1 red king

$= 2$

$$P(A) = \frac{m}{n} = \frac{2}{52}$$

$$P(A) = \frac{1}{26}$$

⑥ Find the probability of getting 2 diamonds if we draw 2 cards at random from a pack of 52 cards.

Sol: Total no. of cards $= 52$

2 cards are picked from 52 cards in $n = {}^{52}C_2$ ways

2 diamonds are picked from 13 cards in $m = {}^{13}C_2$ ways

$$P(A) = \frac{^{13}C_2}{^{52}C_2} = \frac{\frac{13\times12}{\cancel{2}}}{\frac{52\times51}{\cancel{2}}}$$

$$= \frac{\overset{1}{\cancel{13}}\times\overset{3\,1}{\cancel{12}}}{\underset{4\,1}{\cancel{52}}\times\underset{17}{\cancel{51}}}$$

$$P(A) = \frac{1}{17}$$

⑦ 3 cards are drawn from a pack of 52 cards. Find the probability that

i) 3 are spades

ii) 2 spades one diamond

iii) 1 spade, 1 diamond, 1 heart

Sol: 3 cards are picked from 52 cards in $^{52}C_3$ ways

i Getting 3 spades

$$m = {^{13}C_3}$$

$$P(A) = \frac{^{13}C_3}{^{52}C_3} = \frac{\frac{13\times12\times11}{\cancel{3\times2\times1}}}{\frac{52\times51\times50}{\cancel{3\times2\times1}}}$$

$$= \frac{\overset{1}{\cancel{13}}\times\overset{3\,1}{\cancel{12}}\times11}{\underset{4\,1}{\cancel{52}}\times\underset{17}{\cancel{51}}\times50}$$

$$P(A) = \frac{11}{850}$$

(ii) 2 spades and 1 diamond

$m = {}^{13}C_2 \cdot {}^{13}C_1$

$$P(A) = \frac{{}^{13}C_2 \times {}^{13}C_1}{{}^{52}C_3} = \frac{\frac{13\times12}{2}\times 13}{\frac{52\times51\times50}{3\times2\times1}}$$

$$= \frac{\cancel{13}\times\cancel{12}\times 13}{\cancel{2}} \times \frac{3\times\cancel{2}}{\cancel{52}\times\cancel{51}\times 50}$$

$$= \frac{39}{850}$$

(iii) 1 Spade, 1 diamond, 1 heart

$m = {}^{13}C_1 \cdot {}^{13}C_1 \cdot {}^{13}C_1$

$$P(A) = \frac{{}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1}{{}^{52}C_3} = \frac{\cancel{13}\times 13\times 13\times\cancel{3}\times\cancel{2}\times 1}{\cancel{52}\times\cancel{51}\times 50}$$

$$= \frac{169}{1700}$$

(8) What is the probability of drawing ace from a well shuffled pack of 52 playing cards.

Sol:

$n = 52$ cards

$m = 4$ (ace cards)

$$P(A) = \frac{m}{n} = \frac{\cancel{4}^{1}}{\cancel{52}_{13}}$$

$$P(A) = \frac{1}{13}$$

⑨ A bag contains 5 red balls, 8 blue balls, 11 white balls. 3 balls are drawn from a box. Find the probability that

i) 1 Red, 1 Blue and 1 white

ii) 2 whites and 1 Red

iii) 3 white

<u>Sol:</u> 3 balls are picked from 24 balls in $^{24}C_3$ ways

$$n = {}^{24}C_3$$

$\boxed{\text{i}}$ 1 Red, 1 blue & 1 white

$$m = {}^{5}C_1 \times {}^{8}C_1 \times {}^{11}C_1$$

$$P(A) = \frac{{}^{5}C_1 \times {}^{8}C_1 \times {}^{11}C_1}{{}^{24}C_3} = \frac{5\times8\times11}{\frac{24\times23\times22}{3\times2\times1}}$$

$$= \frac{5\times\cancel{8}\times\cancel{11}\times\cancel{3}\times\cancel{2}}{\cancel{24}_{8}\times23\times\cancel{22}}$$

$$P(A) = \frac{5}{23}$$

$\boxed{\text{ii}}$ 2 whites & 1 Red

$$m = {}^{11}C_2 \times {}^{5}C_1$$

$$P(A) = \frac{{}^{11}C_2 \times {}^{5}C_1}{{}^{24}C_3} = \frac{\frac{11\times\cancel{10}^{5}}{\cancel{2}_{1}}\times5}{\frac{24\times23\times22}{3\times2\times1}} = \frac{25\times\cancel{11}\times\cancel{6}}{\cancel{24}_{4}\times23\times\cancel{22}_{2}}$$

$$= \frac{25}{184}$$

$\boxed{iii}$ 3 white

$$P(A) = \frac{{}^{11}C_3}{{}^{24}C_3} = \frac{\frac{11\times10\times9}{3\times2\times1}}{\frac{24\times23\times22}{3\times2\times1}}$$

$$= \frac{\cancel{11}^{1}\times\cancel{10}^{5}\times\cancel{9}^{3}}{\cancel{24}_{8}\times23\times\cancel{22}_{2}}$$

$$= \frac{15}{184}$$

SETS

Set: A collection of well-defined objects is called a set. The objects containing the set, are called elements.

We use capital letters to represent sets and small letters to represent elements.

Sub-Set: ($\subseteq$) Suppose A is a set, B is a set such that $\forall$ every element of B belonging to set A then we say B is a sub-set of A and we writ as $B \subseteq A$.

Union ($\cup$): Let A & B be two sets. Union of A & B is a set of all those elements which are belonging to either A or B or both.

We represent it as $A \cup B$

Intersection ($\cap$): Let A and B be two sets. The intersection of A and B is a set of all those elements which are common to A and B.

We represent it as $A \cap B$

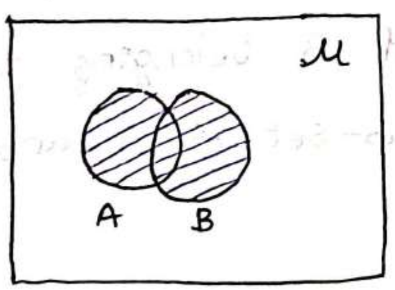
~~Destroying Sets:~~

Disjoint Sets: If $A \cap B = \phi$. That means A and B donot have any element in common. In this case we say that A and B are disjoint.

Complement of A (A^c): The set of elements which don't belong to A.

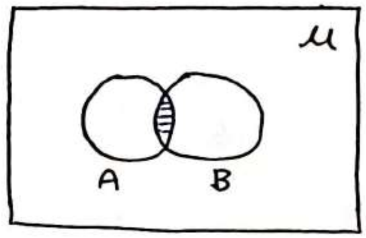
Venn-Diagrams: Rectangle represent universal set. Circles represent sets.

Eg: $A \cup B$



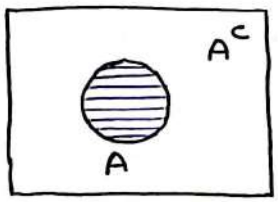
Venn Diagram – $A \cup B$

$A \cap B$



Venn Diagram – $A \cap B$

A^c



Venn Diagram – A^c

<u>Formulae:</u>

* $A \cup A = A$
* $A \cap A = A$
* $A \cup B = B \cup A$
* $(A \cup B) \cup C = A \cup (B \cup C)$
* $A \cap B = B \cap A$
* $A \cup \phi = A$
* $A \cap \phi = \phi$
* $A \cup U = U$
* $A \cap U = A$
* $(A^c)^c = A$
* $A \cap A^c = \phi$

De-Morgan Laws:

* $(A \cap B)^c = A^c \cup B^c$

* $(A \cup B)^c = A^c \cap B^c$

Distributive Laws:

* $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

* $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Axioms of Probability:

The axioms of Probability are

1) $0 \le P(A) \le 1$

2) * $P(S) = 1$

3) * If A and B are any 2 mutually exclusive events then $\boxed{P(A \cup B) = P(A) + P(B)}$

12/09/21

Some elementary theorems:

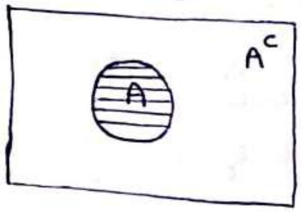
Theorem-1:

Statement → Probability of complementary event is

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

Proof →

* Let A and A^c be any two events



From the figure, we have

$$S = A \cup A^c$$

and also from the figure, A and A^c are mutually exclusive events.

Then By Axiom (3)

$$P(S) = P(A \cup A^c)$$

$$P(S) = P(A) + P(A^c)$$

From axiom(2), $P(S) = 1$

$$1 = P(A) + P(A^c)$$

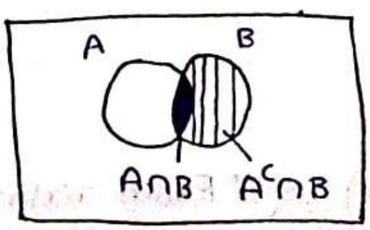
$$\boxed{P(A^c) = 1 - P(A)}$$

Theorem-2:

Statement → For any events A and B

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

Proof → Let A and B be any two events



From the figure, we have

$A \cap B$ & $A^c \cap B$ are mutually exclusive events

Also from figure we have

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

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

$$P(B) = P(A \cap B) + P(A^c \cap B) \quad \text{(From axiom-3)}$$

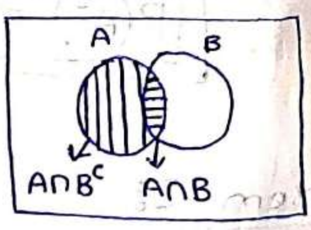
$$\boxed{\boxed{P(A^c \cap B) = P(B) - P(A \cap B)}}$$

Theorem-3:

Statement → If A and B are any two events then

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

Proof → Let A and B be any two events



From the figure, we have

$A \cap B$ and $A \cap B^c$ are mutually exclusive events

Also from figure, we have

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

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

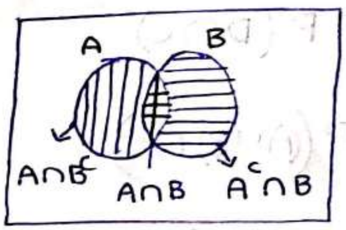
$$P(A) = P(A \cap B^c) + P(A \cap B) \quad (\because \text{From axiom-3})$$

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

ADDITION THEOREM OF PROBABILITY: For any two events A and B

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

Proof → Let A and B be any two events



From the figure we have,

$A\cap B^c, A\cap B, A^c\cap B$ are mutually exclusive events and

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

$$P(A\cup B) = P\left[(A\cap B^c)\cup(A\cap B)\cup(A^c\cap B)\right]$$

$$P(A\cup B) = P(A\cap B^c)+P(A\cap B)+P(A^c\cap B) \quad (\because \text{By axiom } 3)$$

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

$(\because$ From theorem ② & ③$)$

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

MULTIPLICATION THEOREM OF PROBABILITY: For any three events A, B and C

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

Proof → Let A, B and C are three events

Consider

$$P(\underbrace{A\cup B}_{D}\cup C) = P(D\cup C)$$

$$= P(D)+P(C) - P(D\cap C)$$

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

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

$$\boxed{\because (A\cup B)\cap C = (A\cap C)\cup(B\cap C)}$$

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

($\because$ From Addition Theorem)

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

$$\boxed{\therefore P(A\cup B\cup C) = P(A)+P(B)+P(C)-P(A\cap B)-P(B\cap C)-P(C\cap A)+P(A\cap B\cap C)}$$

Formulaes:

* $P(A^c \cap B^c) = P(A \cup B)^c = 1 - P(A \cup B)$

* $P(A^c \cup B^c) = P(A \cap B)^c = 1 - P(A \cap B)$

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

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

* $P(A/B^c) = \dfrac{P(A \cap B^c)}{P(B^c)}$

① If $P(A) = \frac{1}{5}$, $P(B) = \frac{2}{3}$ $P(A \cap B) = \frac{1}{15}$ then find

i) $P(A \cup B)$

ii) $P(A^c \cap B)$

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

iv) $P(A^c \cap B^c)$

v) $P(A^c \cup B^c)$

Sol: Given $P(A) = \frac{1}{5}$

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

$P(A \cap B) = \frac{1}{15}$

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

$= \frac{1}{5} + \frac{2}{3} - \frac{1}{15}$

$= \frac{13}{15} - \frac{1}{15}$

$= \frac{\cancel{12}^{4}}{\cancel{15}_{5}}$

$= \frac{4}{5}$

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

$= \frac{2}{3} - \frac{1}{15}$

$= \frac{\cancel{27}\,\cancel{9}\,3}{\cancel{3}_1 \times \cancel{15}\,5}$

$= \frac{3}{5}$

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

$= \frac{1}{5} - \frac{1}{15}$

$= \frac{\cancel{10}\,2}{\cancel{5}_1 \times 15}$

$= \frac{2}{15}$

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

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

$= 1 - \frac{4}{5}$

$= \frac{1}{5}$

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

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

$= 1 - \frac{1}{15}$

$= \frac{14}{15}$

② If $P(A\cup B)=\frac{4}{5}$, $P(B^c)=\frac{1}{3}$, $P(A\cap B)=\frac{1}{5}$ find

i) $P(B)$

ii) $P(A)$

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

Sol: Given $P(A\cup B)=\frac{4}{5}$

$P(B^c)=\frac{1}{3}$

$P(A\cap B)=\frac{1}{5}$

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

$= 1-\frac{1}{3}$

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

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

$\frac{4}{5} = P(A)+\frac{2}{3}-\frac{1}{5}$

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

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

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

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

$= \frac{2}{3}-\frac{1}{5}$

$= \frac{7}{15}$

14/09/21

③ Two dice are thrown. Let A be the ~~sum of~~ event that sum of points on the faces is 9. Let B be the event that atleast one number is 6. Find the probabilities of the following events.

i) $A \cap B$ ii) $A \cup B$ iii) $A^c \cap B$ iv) $A \cap B^c$ v) $A^c \cap B^c$

vi) $A^c \cup B^c$

Sol: Given two dice are thrown i.e $n = 36$

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

Let A be the event of getting sum as 9

m = {(3,6) (6,3) (4,5) (5,4)}

$m = 4$

$P(A) = \frac{m}{n} = \frac{4}{36} = \frac{1}{9}$

Let B be the event that atleast one number is 6

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

$m = 11$

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

$\boxed{i}$ $P(A\cap B) = \{(3,6)\ (6,3)\}$

$= 2$

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

$\boxed{ii}$ $P(A\cup B) = P(A) + P(B) - P(A\cap B)$

$= \frac{4}{36} + \frac{11}{36} - \frac{2}{36}$

$= \frac{13}{36}$

$\boxed{iii}$ $P(A^c\cap B) = P(B) - P(A\cap B)$

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

$= \frac{9}{36}$

$= \frac{1}{4}$

$\boxed{iv}$ $P(A\cap B^c) = P(A) - P(A\cap B)$

$= \frac{1}{9} - \frac{1}{18}$

$= \frac{1}{18}$

$\boxed{v}$ $P(A^c\cap B^c) = P(A\cup B)^c$

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

$= 1 - \frac{13}{36}$

$= \frac{23}{36}$

$\boxed{vi}$ $P(A^c\cup B^c) = P(A\cap B)^c$

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

$= 1 - \frac{1}{18} = \frac{17}{18}$

(4) If $P(A)=a$, $P(B)=b$, $P(A\cap B)=c$ Express the following Probabilities in terms of a,b,c

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

v) $P(A^c \cup B)$ vi) $P((A\cap B)^c)$ vii) $P((A\cup B)^c)$ viii) $P(A^c\cap(A\cup B))$

ix) $P(A\cup(A^c\cap B))$

Sol: $P(A)=a$ $P(B)=b$ $P(A\cup B)=?$ $P(A\cap B)=c$

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

i) $P(A^c\cup B^c) = P(A\cap B)^c$
$$= 1 - P(A\cap B)$$
$$= 1-c$$

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

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

iv) $P(A^c\cap B^c) = P(A\cup B)^c$
$$= 1 - P(A\cup B)$$
$$= 1 - [a+b-c]$$
$$= c-a-b+1$$

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

$= 1 - P(A) + P(B) - [b - c]$

$= 1 - a + \cancel{b} - \cancel{b} + c$

$= 1 - a + c$

vi) $P((A \cap B)^c) = 1 - P(A \cap B)$

$= 1 - c$

vii) $P((A \cup B)^c) = 1 - P(A \cup B)$

$= 1 - (a + b - c)$

$= c - a - b + 1$

viii) $P(A^c \cap (A \cup B)) = \cancel{\boxed{P(A \cap B)}}$

$P[(A^c \cap A) \cup (A^c \cap B)]$

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

$= 0 + P(A^c \cap B) - 0$

$= b - c$

ix) $P(A \cup (A^c \cap B)) = \cancel{\boxed{P[(A \cup A^c) \cap (A \cup B)]}}$

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

$= a + b - c$

⑤ Among 150 students 80 are studying Maths, 40 are studying Physics and 30 are studying Maths & Physics. If a student is chosen at random, find the probability that the student

a) Studying Maths or Physics

b) Studying Neither maths nor physics

Sol: Let A be the event of student studying Maths

Let B be the event of student studying Physics

$P(A) = \frac{80}{150}$ $P(B) = \frac{40}{150}$ $P(A \cap B) = \frac{30}{150}$

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

$= \frac{80}{150} + \frac{40}{150} - \frac{30}{150}$

$= \frac{\cancel{90}\,3}{\cancel{150}\,5}$

$= \frac{3}{5}$

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

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

$= 1 - \frac{3}{5}$

$= \frac{2}{5}$

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

Sol: Total no. of bolts $= 4+6 = 10$

Probability that 2^{nd} bolt is good if the first bolt is bad is

$$= \frac{{}^{6}C_1}{{}^{10}C_1} \cdot \frac{{}^{4}C_1}{{}^{9}C_1}$$

$$= \frac{\cancel{6}^{2} \times 4}{\cancel{10}_{5} \times \cancel{9}_{3}}$$

$$= \frac{4}{15}$$

⑦ A class has 10 boys and 5 girls. 3 students are selected at random one after the other. Find the probability that

a) First two are boys and third one is girl

b) First and third are of same sex and 2nd is of opposite sex

Sol: Total number of students = 15

a) Probability that first 2 are boys and third is girl $= \frac{{}^{10}C_1}{{}^{15}C_1} \cdot \frac{{}^{9}C_1}{{}^{14}C_1} \cdot \frac{{}^{5}C_1}{{}^{13}C_1}$

$$= \frac{\cancel{10}^{5} \times \cancel{9}^{3} \times \cancel{5}^{1}}{\cancel{15}_{3} \times \cancel{14}_{7} \times 13}$$

$= \frac{15}{91}$

b) Probability that first & third are of same sex and second is of opposite sex

1st possibility

B G B

$$\text{Probability} = \frac{{}^{10}C_1}{{}^{15}C_1} \cdot \frac{{}^{5}C_1}{{}^{14}C_1} \cdot \frac{{}^{9}C_1}{{}^{13}C_1}$$

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

$$= \frac{15}{91}$$

2nd possibility

G B G

$$\text{Probability} = \frac{{}^{5}C_1}{{}^{15}C_1} \cdot \frac{{}^{10}C_1}{{}^{14}C_1} \cdot \frac{{}^{4}C_1}{{}^{13}C_1}$$

$$= \frac{\cancel{5}^{1} \times 10 \times \cancel{4}^{2}}{\cancel{15}_{3} \times \cancel{14}_{7} \times 13}$$

$$= \frac{20}{13 \times 7 \times 3}$$

$$\text{Required Probability} = \frac{15}{13 \times 7} + \frac{20}{13 \times 7 \times 3}$$

$$= \frac{5}{13 \times 7}\left[\frac{3}{1} + \frac{4}{3}\right]$$

$$= \frac{5}{91}\left[\frac{13}{3}\right]$$

$$= \frac{\cancel{65}\,5}{\cancel{273}\,21}$$

$$= \frac{5}{21}$$

15/09/21

Conditional Probability:

Independent Events → Two events are said to be independent if the happening of an event is not affected by the happening of the other event.

In other words, happening of an event doesnot depend upon the happening of the other.

Eg: If we draw a card from a pack of 52 cards and replace it before we draw a second card. The second draw is independent of the first one

Conditional Probability: If B is any arbitrary event in a sample space and $P(B)>0$. The Probability of A is given and B is defined then

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

Similarly,

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

NOTE: 1) From conditional probability, we have

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

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

General Multiplication Rule: If A and B are any events in S then

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

NOTE: If A and B are independent events, then

$$\boxed{P(A \cap B) = P(A) \cdot P(B)}$$

* **Theorem - 1:** If A and B are independent events, then A^c and B^c are also independent events.

Proof: Given A and B are independent events

$$P(A \cap B) = P(A) \cdot P(B) \quad \text{--- (1)}$$

Now we have to prove

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

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

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

$$= 1 - [P(A) + P(B) - P(A \cap B)]$$

$$P(A^c \cap B^c) = 1 - P(A) - P(B) + P(A \cap B)$$

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

$$= 1[1 - P(A)] - P(B)(1 - P(A))$$

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

$$\therefore P(A^c \cap B^c) = P(A^c).P(B^c)$$

* <u>Theorem-2</u>: If A and B are independent events then A and B^c are independent.

<u>Proof</u>: Given A and B are independent events

We have to prove that

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

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

$$= P(A) - P(A)P(B)$$

$$= P(A)[1 - P(B)]$$

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

* <u>Theorem-3</u>: If A, B, C are mutually independent events then $A \cup B$ and C are also independent.

<u>Proof</u>: Given A, B, C are mutually independent events

We have to prove that,

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

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

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

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

$= P(A).P(C) + P(B).P(C) - P(A)P(B)P(C)$

$= P(C)\left[P(A) + P(B) - P(A)P(B)\right]$

$= P(C)\left[P(A) + P(B) - P(A\cap B)\right]$

$\therefore P((A\cup B)\cap C) = P(C)\,P(A\cup B)$

1) The probabilities that students A, B, C, D solve a Problem are $\frac{1}{3}, \frac{2}{5}, \frac{1}{5}$ and $\frac{1}{4}$ respectively. If all of them try to solve the problem. What is the probability that the problem is solved

Sol: Let $P(A)$ = Probability that A solved the problem

$$P(A) = \frac{1}{3} \qquad P(A^c) = \frac{2}{3}$$

Similarly,

$P(B) = \frac{2}{5}$ $\qquad P(B^c) = \frac{3}{5}$

$P(C) = \frac{1}{5}$ $\qquad P(C^c) = \frac{4}{5}$

$P(D) = \frac{1}{4}$ $\qquad P(D^c) = \frac{3}{4}$

$$P(A \cup B \cup C \cup D) = 1 - [P(A \cup B \cup C \cup D)]^c$$

$$= 1 - P(A^c \cap B^c \cap C^c \cap D^c)$$

$$= 1 - P(A^c)\,P(B^c)\,P(C^c)\,P(D^c)$$

$$= 1 - \left(\frac{2}{\cancel{3}}\right)\left(\frac{\cancel{3}}{5}\right)\left(\frac{\cancel{4}}{5}\right)\left(\frac{3}{\cancel{4}}\right)$$

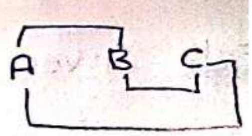
$$= 1 - \frac{6}{25}$$

$$= \frac{19}{25}$$

2) A can hit a target once in 5 shots. B can hit 2 targets in 3 shots. C can hit one target in 4 shots. What is the probability that two shots hit the target

<u>Sol:</u> Let $P(A)$ be the probability that A hit the target is

$$P(A) = \frac{1}{5} \qquad P(A^c) = \frac{4}{5}$$



Similarly

$$P(B) = \frac{2}{3} \qquad P(C) = \frac{1}{4}$$

$$P(B^c) = \frac{1}{3} \qquad P(C^c) = \frac{3}{4}$$

Required Probability

$$= P[(A \cap B \cap C^c) \cup (A \cap B^c \cap C) \cup (A^c \cap B \cap C)]$$

$$= P(A \cap B \cap C^c) + P(A \cap B^c \cap C) + P(A^c \cap B \cap C)$$

$= P(A).P(B).P(C^c) + P(A)P(B^c)P(C) + P(A^c)P(B)P(C)$

$= \left(\frac{1}{5}\right)\left(\frac{2}{3}\right)\left(\frac{3}{4}\right) + \frac{1}{5}\left(\frac{1}{3}\right)\left(\frac{1}{4}\right) + \left(\frac{4}{5}\right)\left(\frac{2}{3}\right)\left(\frac{1}{4}\right)$

$= \dfrac{6+1+8}{5\times3\times4}$

$= \dfrac{\cancel{15}}{\cancel{15}\times4}$

$= \frac{1}{4}$

3) The probabilities of passing in subjects A, B, C, D are $\frac{3}{4}, \frac{2}{3}, \frac{4}{5}$ and $\frac{1}{2}$. To qualify in the examination a student should pass in A and two subjects among the three. What is the probability of qualifying in the exam.

Sol: Given $P(A) = \frac{3}{4}$ $\quad P(A^c) = \frac{1}{4}$

$P(B) = \frac{2}{3}$ $\quad P(B^c) = \frac{1}{3}$

$P(C) = \frac{4}{5}$ $\quad P(C^c) = \frac{1}{5}$

$P(D) = \frac{1}{2}$ $\quad P(D^c) = \frac{1}{2}$

There are four possibilities to qualify the exam

1) To pass in A,B,C and fail in D 1) A B C D^c

2) To pass in A,C,D and fail in B 2) A C D B^c

3) To pass in A,B,D and fail in C 3) A B D C^c

4) To pass in A,B,C,D 4) A B C D

Required Probability $= P\left[(A\cap B\cap C\cap D^c)\cup(A\cap C\cap D\cap B^c)\cup(A\cap B\cap D\cap C^c)\cup(A\cap B\cap C\cap D)\right]$

$= P(A\cap B\cap C\cap D^c)+P(A\cap C\cap D\cap B^c)+P(A\cap B\cap D\cap C^c)+P(A\cap B\cap C\cap D)$

$= P(A)P(B)P(C)P(D^c)+P(A)P(C)P(D)P(B^c)+P(A)P(B)P(D)P(C^c)+P(A)P(B)P(C)P(D)$

$= \frac{3}{4}\left(\frac{2}{3}\right)\left(\frac{4}{5}\right)\left(\frac{1}{2}\right)+\frac{3}{4}\left(\frac{4}{5}\right)\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)+\left(\frac{3}{4}\right)\left(\frac{2}{3}\right)\left(\frac{1}{2}\right)\left(\frac{1}{5}\right)+\left(\frac{3}{4}\right)\left(\frac{2}{3}\right)\left(\frac{4}{5}\right)\left(\frac{1}{2}\right)$

$= \frac{24+12+6+24}{120}$

$= \frac{\cancel{66}^{11}}{\cancel{120}_{20}}$

$= \frac{11}{20}$

4) Determine

a) $P(B/A)$ b) $P(A/B^c)$ if A and B are events with

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

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

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

$P(A\cap B) = \frac{7}{12} - \frac{1}{2}$

$P(A\cap B) = \frac{1}{12}$

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

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

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

$= \frac{1}{4}$

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

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

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

$= \frac{\frac{1}{4}}{\frac{3}{4}}$

$= \frac{1}{4} \times \frac{4}{3}$

$= \frac{1}{3}$

***

Baye's Theorem: Suppose $B_1, B_2, \ldots B_n$ are mutually exclusive events of a sample space S such that $P(B_i) > 0$ $i = 1,2,3 \ldots n$ and A is any arbitrary event of S such that $P(A) > 0$ and $A \subseteq \bigcup_{i=1}^{n} B_i$ $i=1,2 \ldots n$ then the conditional Probability of B_i given $P(A)$ for $i=1,2 \ldots n$ is equal to

$$\boxed{P(B_i/A) = \frac{P(B_i)\,P(A/B_i)}{\sum_{i=1}^{n} P(B_i)\,P(A/B_i)}}$$

Proof: Suppose B_1, B_2, B_3 B_n are the set of exhaustive and mutually exclusive events.

Let A is any arbitrary event of S

We Know That, $P(A \cap B_i) = P(B_i)\,P\left(\frac{A}{B_i}\right) = P(A)\,P(B_i/A)$ —— (1) —— (2)

By definition of Conditional Probability, we have

$$P(B_i/A) = \frac{P(A \cap B_i)}{P(A)} \quad \text{—— ②}$$

Sub ① in ②

$$P(B_i/A) = \frac{P(B_i)\,P(A|B_i)}{P(A)} \quad \text{—— ③}$$

Given $A \subseteq \bigcup_{i=1}^{n} B_i$

From the figure we have

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \quad \cup (A \cap B_n)$$

$$P(A) = P\left[(A \cap B_1) \cup (A \cap B_2) \cup \quad \cup (A \cap B_n)\right]$$

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n)$$

$P(A) = P(B_1 \cap A) + P(B_2 \cap A) + \dots + P(B_n \cap A)$

$P(A) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + \dots + P(B_n)P(A/B_n)$ (By Definition of Conditional Probability)

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A/B_i) \quad (4)$$

Sub $P(A)$ in (3)

$$\boxed{P(B_i/A) = \frac{P(B_i)P(A/B_i)}{\sum_{i=1}^{n} P(B_i)P(A/B_i)}}$$

~~Sub (1)~~

20/09/21

1) In a bolt factory machines A, B, C manufacture 20%, 30%, 50% of the total of their output 6%, 3% and 2% are defective. A bolt is drawn (→ Denominator) at random and found to be defective. What is the probability that is manufactured by machines A, B and C.

Sol: Let A be the probability that the bolt was manufactured by machine A

$P(A) = 20\% = \frac{20}{100} = 0.2$

$P(B) = 30\% = \frac{30}{100} = 0.3$

$P(C) = 50\% = \frac{50}{100} = 0.5$

Let D be the probability of the defective bolt manufactured by machine A, B & C

$P(D/A) = 6\%$

$P(D/B) = 3\%$

$P(D/c) = 2\%$

By Baye's Theorem,

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

$$= \frac{0.2\,(0.06)}{0.2(0.06) + (0.3)(0.03) + (0.5)(0.02)}$$

$$= \frac{0.012}{0.012 + 0.009 + 0.01}$$

$$= \frac{0,012}{0,031}$$

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

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

$$= \frac{(0.3)\,(0.03)}{(0.2)(0.06) + (0.3)(0.03) + (0.5)(0.02)}$$

$$= \frac{0,009}{0,031}$$

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

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

$$= \frac{0.5(0.02)}{(0.2)(0.06)+(0.3)(0.03)+(0.5)(0.02)}$$

$$= \frac{0.01 \times 10^3}{0.031 \times 10^3}$$

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

2) There are three boxes (I, II, III). Box I contains 4 red, 5 blue & 6 white balls. Box II contains 3 red, 4 blue & 5 white balls. Box III contains 5 red, 10 blue & 5 white balls. One box is chosen and one ball is drawn from it. What is the probability that

a) The red ball is drawn from Box I

b) The blue ball is drawn from Box II

c) The white ball is drawn from Box III

Sol: Let B_1, B_2, B_3 are the three boxes i.e

$P(B_1) = \frac{1}{3}$

$P(B_2) = \frac{1}{3}$

$P(B_3) = \frac{1}{3}$

$P(R/B_1) = \frac{4}{15}$ $\quad P(B/B_1) = \frac{5}{15}$ $\quad P(W/B_1) = \frac{6}{15}$

$P(R/B_2) = \frac{3}{12}$ $\quad P(B/B_2) = \frac{4}{12}$ $\quad P(W/B_2) = \frac{5}{12}$

$P(R/B_3) = \frac{5}{20}$ $\quad P(B/B_3) = \frac{10}{20}$ $\quad P(W/B_3) = \frac{5}{20}$

By Baye's Theorem

a) Probability of getting a red ball from Box-I

$$P(B_1/R) = \frac{P(B_1)\,P(R/B_1)}{P(B_1)\,P(R/B_1) + P(B_2)P(R/B_2) + P(B_3)\,P(R/B_3)}$$

$$= \frac{\cancel{\frac{1}{3}}\left(\frac{4}{15}\right)}{\cancel{\frac{1}{3}}\left[\frac{4}{15} + \frac{\cancel{3}^{1}}{\cancel{12}_{4}} + \frac{\cancel{5}^{1}}{\cancel{20}_{4}}\right]}$$

$$= \frac{\frac{4}{15}}{\frac{4}{15} + \frac{1}{2}}$$

$$= \frac{\frac{4}{15}}{\frac{23}{30}}$$

$$= \frac{4}{\cancel{15}_{1}} \times \frac{\cancel{30}^{2}}{23}$$

$$= \frac{8}{23}$$

b) Probability of getting a blue ball from Box-II

$$P(B_2/B) = \frac{P(B_2)\,P(B/B_2)}{P(B_1)P(B/B_1) + P(B_2)P(B/B_2) + P(B_3)P(B/B_3)}$$

$$= \frac{\cancel{\frac{1}{3}}\left(\frac{1}{3}\right)}{\cdots}$$

$$= \frac{\frac{1}{3}}{\frac{2}{3}+\frac{1}{2}}$$

$$= \frac{\frac{1}{3}}{\frac{7}{6}}$$

$$= \frac{1}{\cancel{3}_{1}} \times \frac{\cancel{6}^{2}}{7}$$

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

c) Probability of getting white ball from Box – III

$$P(B_3/W) = \frac{P(B_3)P(W/B_3)}{P(B_1)P(W/B_1)+P(B_2)P(W/B_2)+P(B_3)P(W/B_3)}$$

$$= \frac{\cancel{\frac{1}{3}}\left(\frac{1}{4}\right)}{\cancel{\frac{1}{3}}\left[\frac{2}{5}+\frac{5}{12}+\frac{1}{4}\right]}$$

$$= \frac{\frac{1}{4}}{\frac{2}{5}+\frac{\cancel{8}^{2}}{\cancel{12}_{3}}}$$

$$= \frac{\frac{1}{4}}{\frac{16}{15}}$$

$$= \frac{1}{4} \times \frac{15}{16}$$

$$= \frac{15}{64}$$

3) In a class, 2% of boys and 3% of girls are having blue eyes. There are 30% girls in the class. If a student is selected and having blue eyes, What is the probability that the student is a girl.

Sol: Let G be the probability of girls i.e

$$P(G) = 30\% = \frac{30}{100} = 0.3$$

Let B be the probability of boys

$$P(B) = 70\% = \frac{70}{100} = 0.7$$

$$P(B.E/G) = 3\% = \frac{3}{100} = 0.03$$

$$P(B.E/B) = 2\% = \frac{2}{100} = 0.02$$

By Baye's Theorem

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

$$= \frac{0.3(0.03)}{0.3(0.03) + (0.7)(0.02)}$$

$$= \frac{\frac{9}{\cancel{1000}}}{\frac{9+14}{\cancel{1000}}}$$

$$= \frac{9}{23}$$

4) In a certain college, 25% of boys & 10% of girls are studying Mathematics. The girls constitute 60% of the students. If a student is selected & is found to be studying Mathematics. Find the probability that the student

a) is a girl b) is a boy

Sol: $P(B) = 40\% = \frac{40}{100} = 0.4$

$P(G) = 60\% = \frac{60}{100} = 0.6$

$P(M/B) = \frac{25}{100} = 0.25$

$P(M/G) = \frac{1\cancel{0}}{10\cancel{0}} = 0.1$

By Baye's Theorem

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

$$= \frac{0.4(0.25)}{0.4(0.25) + (0.6)(0.1)}$$

$$= \frac{0.1}{0.1+0.06} = \frac{0.1 \times 10\cancel{0}}{0.16 \times 1\cancel{0}} = \frac{10}{16}$$

$$= \frac{5}{8}$$

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

$$= \frac{0.06}{0.16}$$

$$= \frac{6}{16}$$

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

21/09/21

5) A businessman goes to hotels X, Y, Z 20%, 50% and 30% of the time respectively. It is known that 5%, 4%, 8% of the rooms in X, Y, Z hotels have faulty plumbing. What is the probability that the businessman room having faulty plumbing is assigned to hotel Z.

Sol: X, Y, Z are the 3 given hotels

$P(X) = 20\% = \frac{20}{100} = 0.2$

$P(Y) = 50\% = \frac{50}{100} = 0.5$

$P(Z) = 30\% = \frac{30}{100} = 0.3$

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

$P(F.P/Y) = 4\% = 0.04$

$P(F.P/Z) = 8\% = 0.08$

By Baye's Theorem,

$$P(Z/F.P) = \frac{P(Z)P(F.P/Z)}{P(X)P(F.P/X)+P(Y)P(F.P/Y)+P(Z)P(F.P/Z)}$$

$$= \frac{(0.3)(0.08)}{(0.2)(0.05)+(0.5)(0.04)+(0.3)(0.08)}$$

$$= \frac{\frac{24}{\cancel{1000}}}{\frac{54}{\cancel{1000}}}$$

$$= \frac{\cancel{24}^{12^{4}}}{\cancel{54}_{27\,9}}$$

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

6) A Box 1 contains 11 cards numbered 1 to 11, Box 2 contains 7 cards numbered 1 to 7. A box is selected at random and a card is drawn. If the no. is even, find the probability that the card is from Box 1.

Sol: Let B_1, B_2 are two boxes

$P(B_1) = \frac{1}{2}$

$P(B_2) = \frac{1}{2}$

Let E_1 be the event of getting even number from box 1 i.e $P(E_1/B_1) = \frac{5}{11}$

Let E_2 be the event of getting even number from box 2 i.e $P(E_2/B_2) = \frac{3}{7}$

By Baye's Theorem,

$$P\left(B_1/E\right) = \frac{P(B_1)P(E_1/B_1)}{P(B_1)P(E_1/B_1) + P(B_2)P(E_2/B_2)}$$

$$= \frac{\cancel{\tfrac{1}{2}}\left[\frac{5}{11}\right]}{\cancel{\tfrac{1}{2}}\left(\frac{5}{11} + \frac{3}{7}\right)}$$

$$= \frac{\frac{5}{11}}{\frac{68}{77}}$$

$$= \frac{5}{\cancel{11}_{1}} \times \frac{\cancel{77}^{7}}{68}$$

$$= \frac{35}{68}$$

21/09/21

Chapter-2 RANDOM VARIABLES

Random Variables: A real variable X whose value is determined by the outcome of a random experiment is called a random variable.

* A random variable X can also be regarded as a real value function defined on the sample space S of a random experiment such that for each point x of the sample space, $f(x)$ is the probability of occurence of the event represented by x.

Eg: When 2 coins are tossed,

Let E be the event of getting a tail

The Sample Space $S = \{HH, HT, TH, TT\}$

The probability of getting no tail $= 0$

The probability of getting one tail $= 1$

The probability of getting two tails $= 2$

$\therefore$ The random variable $X = \{0, 1, 2\}$

* Random Variables are of two types:

1) Discrete

2) Continuous

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

Eg: Tossing a coin, Throwing a die etc

Continuous Random Variable: A random variable X which can take values continuously i.e which takes all possible values in a given interval is called Continuous Random Variable.

Eg: The height, weight and age of an individual; Temperature and Time etc

Probability Distribution Function: Let X be a random variable. Then the Probability Distribution Function assosciated with X is defined as the probability that the outcome of an experiment will be one of the outcomes. The function $F(x)$ is defined by

$$\boxed{F_X(x) = P(X \leq x)}$$

Discrete Probability Distribution: Probability Distribution of a random Variable is the set of its Possible values together with their respective Probabilities, i.e

1) $P(x_i) \geq 0$ 2) $\sum_{i=1}^{n} P_i = 1$

* If the function P satisfies above 2 conditions then the function P is called Discrete Probability Distribution.

Probability Density Function: The Probability Density Function is defined as the derivative of the Probability Distribution Function of the random Variable X i.e

$$f_X(x) = \frac{d}{dx}\left[F_X(i)\right]$$

Expectation (or) Mean (or) Expected Value of x: Suppose a random Variable X assumes the values $x_1, x_2 \dots x_n$ w.r.to the probabilities $P_1, P_2 \dots P_n$. Then the mean (or) expectation is denoted by $E(X)$ or μ and is defined as The sum of products of different Values of x and corresponding probabilities i.e

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

22/09/21

SAQ

Theorem: If X is a random Variable & K is a constant then $\boxed{E(X+K) = E(X) + K}$

Proof:

Given X is a random Variable

By definition of expectation, we have $E(x) = \sum_{i=1}^{n} x_i P_i$

Consider
$$E(X+K) = \sum_{i=1}^{n} (x_i + k) P_i$$
$$= \sum_{i=1}^{n} x_i P_i + \sum_{i=1}^{n} K P_i$$
$$= \sum_{i=1}^{n} x_i P_i + K \sum_{i=1}^{n} P_i$$
$$= E(X) + K(1) \qquad \left(\because \sum_{i=1}^{n} P_i = 1\right)$$
$$E(X+K) = E(X) + K$$

Theorem: If X is a random variable then
$$\boxed{E(KX) = K.E(X)}$$

Proof:

Given X is a random Variable

By definition of expectation, we have
$$E(X) = \sum_{i} x_i P_i$$

Consider $E(KX) = \sum_{i=1}^{n} K x_i P_i$

$$= K \sum_{i=1}^{n} x_i P_i$$

$$E(KX) = K.E(X)$$

<u>Theorem</u>: If X is a random variable and a, b are constants then $\boxed{E(aX+b) = a(E(X)) + b}$

<u>Proof</u>:

Given X is the random variable and a, b are constants

By definition of expectation, $E(x) = \sum_{i=1}^{n} x_i P_i$

Consider $E(aX+b) = \sum_{i=1}^{n} (ax_i+b) P_i$

$$= a \sum_{i=1}^{n} x_i P_i + b \sum_{i=1}^{n} P_i$$

$$= a\,E(x) + b(1)$$

$$E(ax+b) = a\,E(x) + b$$

<u>Theorem</u>: If X and Y are two discrete random variables then $\boxed{E(X+Y) = E(X) + E(Y)}$

Proof: Given X, Y are discrete random variables

Consider

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

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

$$= \sum_{i=1}^{n} x_i \sum_{j=1}^{n} P_i \cdot P_j + \sum_{j=1}^{n} y_j \sum_{i=1}^{n} P_i P_j$$

$$= \sum_{i=1}^{n} x_i P_i \sum_{j=1}^{n} P_j + \sum_{j=1}^{n} y_j P_j \sum_{i=1}^{n} P_i$$

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

$$E(X+Y) = E(X) + E(Y)$$

Variance: The variance of a discrete random variable is denoted by "σ^2" and is defined as

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

Standard Deviation (S.D): The standard deviation is the square root of the variance.

$$\boxed{S.D = \sqrt{\sigma^2}}$$

Theorem: The variance (σ^2) of the discrete random Variable is

$$\boxed{\sigma^2 = E(x^2) - [E(x)]^2}$$

Proof: By definition of Variance,

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

$$= E[x - \mu]^2$$

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

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

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

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

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

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

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

Theorem: If x is a discrete random Variable then

$$\boxed{V(ax+b) = a^2 V(x)}$$ where $V(x)$ is Variance of x

and a, b are constants

Proof: Given X is a random Variable

Let $Y = aX + b$ —①

$E(Y) = E(ax + b)$

$E(Y) = a(E(X)) + b$ —②

① − ② $\Rightarrow Y - E(Y) = a[X - E(X)]$

Squaring on Both Sides and taking expectation

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

$Var(Y) = a^2 Var(X)$ ($\because$ By definition of Variance)

$V(Y) = a^2 V(X)$

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

NOTE: While solving the problems, we use the following formulae

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

2) Variance $\sigma^2 = \sum_{i=1}^{n} x_i^2 P_i - \mu^2$

3) Standard Deviation $= \sqrt{\text{Variance}}$

1) A random Variable X has the following probability distribution

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

Determine a) K

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

c) Mean

d) Variance

e) If $P(X\leq K) > \frac{1}{2}$ find minimum Value of K

f) Determine the distribution function of X

Sol:

a) To find K

WKT, Sum of Probabilities = 1

$$\sum_{i=1}^{n} P_i = 1$$

$$\sum_{x=0}^{7} P_i = 1 \Rightarrow 0 + K + 2K + 2K + 3K + K^2 + 2K^2 + 7K^2 + K =$$

$$\Rightarrow 10K^2 + 9K = 1$$

$$\Rightarrow 10K^2 + 9K - 1 = 0$$

$$K = \frac{-9 \pm \sqrt{81+40}}{2(10)}$$

$$= \frac{-9 \pm 11}{20}$$

$$= -1, \frac{1}{10}$$

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

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

$= 0 + K + 2K + 2K + 3K + K^2$

$= 8K + K^2$

$= 8\left(\frac{1}{10}\right) + \frac{1}{100}$

$= \frac{81}{100}$

$= 0.81$

$P(x \geq 6) = P(x=6) + P(x=7)$

$= 2K^2 + 7K^2 + K$

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

$= \frac{19}{100}$

$= 0.19$

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

$= K + 2K + 2K + 3K$

$= 8K$

$= 8\left(\frac{1}{10}\right)$

$= 0.8$

c) Mean $= \sum_{i=1}^{n} x_i P_i$

$= 0 + K + 4K + 6K + 12K + 5K^2 + 12K^2 + 49K^2 + 7K$

$= 66K^2 + 30K$

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

$= \frac{366}{100}$

$\mu = 3.66$

d) Variance $= \sum_{i=1}^{n} x_i^2 P_i - \mu^2$

$= [0 + K + 8K + 18K + 48K + 25K^2 + 72K^2 + 343K^2 + 49K] - (3.66)^2$

$= 440K^2 + 124K - (13.3956)$

$= 44\cancel{0}\left(\frac{1}{10\cancel{0}}\right) + 124\left(\frac{1}{10}\right) - 13.3956$

$= 16.8 - 13.3956$

$= 3.4044$

e) $P(x \le 1) = P(x=0) + P(x=1) = 0 + K = \frac{1}{10} = 0.1 < \frac{1}{2}$

$P(x \le 2) = P(x=0) + P(x=1) + P(x=2) = K + 2K = \frac{3}{10} = 0.3 < \frac{1}{2}$

$P(X \le 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3) = K + 2K + 2K = \frac{5}{10} = 0.5 <$

$P(x \le 4) = P(x=0) + P(x=1) + P(x=2) + P(x=3) + P(x=4) = 8K = \frac{9}{10} = 0.8 < \frac{1}{2}$ (Fals

$0.8 > \frac{1}{2}$

$\therefore$ Minimum Value of $K = 4$

f) The distribution function of X is given by the following table

X	$P(x) = P(X \le x)$
0	0
1	$K = \frac{1}{10}$
2	$K + 2K = \frac{3}{10}$
3	$5K = \frac{5}{10}$
4	$8K = \frac{8}{10}$
5	$K^2 + 8K = \frac{1}{100} + \frac{8}{10} = \frac{81}{100}$
6	$3K^2 + 8K = \frac{3}{100} + \frac{8}{10} = \frac{83}{100}$
7	$10K^2 + 9K = 1\cancel{0}\left(\frac{1}{10\cancel{0}}\right) + \frac{9}{10} = \frac{10}{10} = 1$

2) For the discrete probability distribution

x	0	1	2	3	4	5	6
P	0	$2K$	$2K$	$3K$	K^2	$2K^2$	$7K^2+K$

a) Find K

b) Find Mean

c) Find Variance

Sol:-

a) $\sum_{i=1}^{n} P_i = 1$

$\Rightarrow 2K + 2K + 3K + K^2 + 2K^2 + 7K^2 + K = 1$

$10K^2 + 8K - 1 = 0$

$K = \dfrac{-8 \pm \sqrt{64 - 4(10)(-1)}}{2(10)}$

$= \dfrac{-8 \pm \sqrt{104}}{20}$

$= \dfrac{-8 \pm 10.19803903}{20}$

$K = 0.1099$

b) Mean $\mu = \sum_{i=1}^{n} x_i P_i$

$= 0 + 2K + 4K + 9K + 4K^2 + 10K^2 + 42K^2 + 6K$

$= 56K^2 + 21K$

$= 56(0.1099)^2 + 21(0.1099)$

$= 0.6763 + 2.3079$

$= 2.9842$

c) Variance $= \sum_{i=1}^{n} x_i^2 P_i - \mu^2$

$= 0 + 2K + 8K + 27K + 16K^2 + 50K^2 + 252K^2 + 36K - (2.9842)^2$

$= 318K^2 + 73K - 8.9054$

$= 318(0.1099)^2 + 73(0.1099) - 8.9054$

$= 3.8408 + 8.0227 - 8.9054$

$\sigma^2 = 2.9581$

***

3) Let X denote the no. of heads in a single toss of 4 ~~points~~ coins. Determine

a) $P(X<2)$ b) $P(1<X\leq 3)$

Sol: Given 4 coins are tossed

Total no. of cases $= 2^4 = 16$

HHHH	HHHT	HHTT	HTTT	TTTT
	THHH	TTHH	TTTH	
	HTHH	HTHT	TTHT	
	HHTH	THTH	THTT	
		HTTH		
		THHT		

Let E be the event of getting heads

x	0	1	2	3	4
$P(x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

a) $P(X<2)$

$= P(X=0) + P(X=1)$

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

$= \frac{5}{16}$

b) $P(1<X\leq 3)$

$= P(X=2)+P(X=3)$

$= \frac{6}{16} + \frac{4}{16}$

$= \frac{10}{16} = \frac{5}{8}$

***

4) Two dice are thrown. Let X assign to each point (a,b) in S. The maximum of its numbers i.e $X(a,b) = \max(a,b)$. Find the probability distribution, the mean and variance of the distribution.

Sol: Two dice are thrown

Total no. of cases $= 6^2 = 36$

Let E be the event of getting $X(a,b) = \max(a,b)$

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

For maximum '1' favourable cases are $(1,1) = \frac{1}{36}$

For maximum '2' favourable cases are $(1,2)(2,1)(2,2) = \frac{3}{36}$

For maximum '3' favourable cases are $(1,3)\ (2,3)\ (3,3)$
$(3,1)\ (3,2) = \frac{5}{36}$

Probability distribution function

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

a) Mean $(\mu) = \sum_{i=1}^{6} x_i P_i$

$$= \frac{1}{36} + \frac{6}{36} + \frac{15}{36} + \frac{28}{36} + \frac{45}{36} + \frac{66}{36} = \frac{161}{36}$$

$$= 4.472$$

b) Variance $(\sigma^2) = \sum_{i=1}^{6} x_i^2 P_i - \mu^2$

$$= \frac{1}{36} + \frac{12}{36} + \frac{45}{36} + \frac{112}{36} + \frac{225}{36} + \frac{396}{36} - (4.472)^2$$

$$= \frac{791}{36} - (4.472)^2$$

$$= 21.97 - 19.99$$

$$= 1.98$$

5) Let X denote to each point (a, b) in S. The min of its numbers i.e $X(a,b) = \min(a,b)$. Find the Probability distribution, mean & variance of the distribution

Sol: For minimum '1' favourable cases are

(1,1) (1,2) (1,3) (1,4) (1,5) (1,6) $= \frac{11}{36}$

(2,1) (3,1) (4,1) (5,1) (6,1)

For minimum '2' favourable cases are $= \frac{9}{36}$

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

i) Mean $\mu = \sum_{i=1}^{6} P_i x_i$

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

$$= \frac{91}{36}$$

$$= 2.527$$

ii) Variance $\sigma^2 = \sum_{i=1}^{6} P_i x_i^2 - \mu^2$

$$= \frac{11}{36} + \frac{36}{36} + \frac{63}{36} + \frac{80}{36} + \frac{75}{36} + \frac{36}{36} - (2.527)^2$$

$$= 8.3611 - 6.3857$$

$$= 1.975$$

6) Given that $f(x) = \frac{K}{2x}$ is a probability distribution for a random Variable X that can take the Values $x = 0, 1, 2, 3, 4$

a) Find K

b) Mean & Variance of x

Sol: Given $f(x) = \frac{K}{2x}$ $\quad x = 0,1,2,3,4$

x	1	2	3	4
$f(x)$	$\frac{K}{2}$	$\frac{K}{4}$	$\frac{K}{6}$	$\frac{K}{8}$

a) $\sum_{i=1}^{n} P_i = 1$

$\Rightarrow \frac{K}{2} + \frac{K}{4} + \frac{K}{6} + \frac{K}{8} = 1$

$\Rightarrow \frac{3K \times 6}{4 \times 6} + \frac{\cancel{14}^{7} K}{\cancel{48}_{24}} = 1$

$\Rightarrow 25K = 24 \Rightarrow K = \frac{24}{25} = 0.96$

b) Mean $\mu = \sum_{i=1}^{4} x_i P_i$

$= \frac{K}{2} + \frac{2K}{4} + \frac{3K}{6} + \frac{4K}{8}$

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

$= 2K$

$= 2\left(\frac{24}{25}\right)$

$\mu = 1.92$

Variance $\sigma^2 = \sum_{i=1}^{4} x_i^2 P_i - \mu^2$

$= \frac{K}{2} + \frac{\cancel{4}K}{\cancel{4}} + \frac{\cancel{9}^{3}K}{\cancel{6}_{2}} + \frac{16K}{8} - (1.92)^2$

$= 3K + 2K - (1.92)^2$

$= 5(0.96) - (1.92)^2$

$= 4.8 - 3.6864$

$= \ldots 136$

7) A random Variable X is defined as the sum of the numbers on the faces when two dice are thrown. Find the mean.

Sol: Total no. of cases $= 6^2 = 36$

x_i	2	3	4	5	6	7	8	9	10	11	12
$P(x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$$\text{Mean}(\mu) = \sum_{i=2}^{12} P_i x_i$$

$$= 2\left(\frac{1}{36}\right) + 3\left(\frac{2}{36}\right) + 4\left(\frac{3}{36}\right) + 5\left(\frac{4}{36}\right) + 6\left(\frac{5}{36}\right) + 7\left(\frac{6}{36}\right) + 8\left(\frac{5}{36}\right) + 9\left(\frac{4}{36}\right) + 10\left(\frac{3}{36}\right) + 11\left(\frac{2}{36}\right) + 12\left(\frac{1}{36}\right)$$

$$= \frac{2+6+12+20+30+42+40+36+30+22+12}{36}$$

$$= \frac{\cancel{252}^{\cancel{42}\ 7}}{\cancel{36}\ \cancel{6}\ 1}$$

$\mu = 7$

$\boxed{27/09/21}$

8) ~~Given that $f(x) =$~~ Find the mean and Variance of the uniform probability distribution given by

$F(x) = \frac{1}{n}$ For $x = 1, 2, 3 \ldots n$

Sol: The probability distribution is

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

Mean $\mu = \sum_{i=1}^{n} x_i p_i$

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

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

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

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

Variance $\sigma^2 = \sum_{i=1}^{n} x_i^2 p_i - \mu^2$

$$= 1^2\left(\frac{1}{n}\right) + 2^2\left(\frac{1}{n}\right) + 3^2\left(\frac{1}{n}\right) + \dots + n^2\left(\frac{1}{n}\right) - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{1}{n}\left(1^2+2^2+3^2+\dots+n^2\right) - \left(\frac{n+1}{2}\right)^2$$

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

$$= \frac{n+1}{2}\left[\frac{2n+1}{3} - \frac{n+1}{2}\right]$$

$$= \frac{n+1}{2}\left[\frac{2(2n+1)-3n-3}{6}\right]$$

$$= \frac{n+1}{2}\left(\frac{4n+2-3n-3}{6}\right)$$

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

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

9) A sample of 4 items are selected at random from box containing 12 items of which 5 are defective. Find the expected no. of defective items.

Sol: Let X denote number of defective items,

Among 4 items drawn from 12 items

Let X can take the values 0, 1, 2, 3, 4

Given total no. of items = 12

No. of defective items = 5

No. of good items = 12 − 5 = 7.

$P(x=0) = P$ (no defective items)

$$= \frac{{}^7C_4}{{}^{12}C_4}$$

$$= \frac{7 \times \cancel{6} \times \cancel{5} \times \cancel{4}}{\cancel{12} \times 11 \times \cancel{10} \times 9}$$

$$= \frac{7}{99}$$

$P(x=1) = P$ (1 defective & 3 good items)

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

$$= \frac{5 \times 7 \times \cancel{6} \times \cancel{5}}{\cancel{4} \cdot \cancel{12} \times 11 \times \cancel{10} \times 9} \frac{\times\, 3 \times 2 \times 1 \times \cancel{4}}{\times\, 3 \times 2 \times 1} = \frac{35}{99}$$

$P(x=2) = P(2 \text{ defective } \& \text{ 2 good items})$

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

$$= \frac{\frac{5\times \cancel{4}^{2}}{\cancel{2}\cdot 1} \times \frac{7 \times \cancel{6}^{3}}{\cancel{2}\cdot 1} \times 4 \times 3 \times 2 \times 1}{12 \times 11 \times 10 \times 9}$$

$$= \frac{\cancel{10} \times \cancel{21}^{7} \times \cancel{12} \times 2}{\cancel{12} \times 11 \times \cancel{10} \times \cancel{9}\,3}$$

$$= \frac{14}{33}$$

$P(x=3) = P(3 \text{ defective } \& \text{ 1 good item})$

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

$$= \frac{\frac{\cancel{5}^{1} \times \cancel{4}^{2}}{\cancel{2}} \times 7 \times \cancel{4 \times 3 \times 2} \times 1}{\cancel{12} \times 11 \times \cancel{10} \times 9}$$

$$= \frac{14}{99}$$

$P(x=4) = P(\text{all defective items})$

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

$$= \frac{\cancel{5} \times \cancel{4} \times \cancel{3}^{1} \times \cancel{2}^{1} \times 1}{\cancel{12}_{4} \times 11 \times \cancel{10}_{5} \times 9}$$

$$= \frac{1}{99}$$

The probability distribution table

x	0	1	2	3	4
$P(x_i)$	$\frac{7}{99}$	$\frac{35}{99}$	$\frac{14}{33}$	$\frac{14}{99}$	$\frac{1}{99}$

Mean $\mu = \sum_{i=1}^{4} x_i P_i$

$$= 0 + \frac{35}{99} + \frac{28}{33} + \frac{42}{99} + \frac{4}{99}$$

$$= \frac{81}{99} + \frac{28}{33}$$

$$= \frac{81+84}{99}$$

$$= \frac{\cancel{165}\,\cancel{55}\,5}{\cancel{99}\,\cancel{33}\,3}$$

$$= \frac{5}{3} \text{ (or) } \frac{165}{99}$$

P(10) A player tosses 3 fair coins. He wins ₹ 500 if 3 heads appear, ₹ 300 if 2 heads appear, ₹ 100 if 1 head appear. On the other hand, he loses ₹ 1500 if 3 tails appear. Find the expected Gain of the Player.

Sol: Given 3 coins are tossed

Total no. of cases $= 2^3 = 8$

$S = \{HHH, HTT, HTH, HHT, TTT, THH, THT, TTH\}$

Let X denote the gain. The range of $X = \{-1500, 100, 300, 500\}$

The probability of getting 3 heads is

$P(\text{getting } 500\text{ Rs}) = \frac{1}{8}$

The probability of getting 2 heads is

$P(\text{getting Rs } 300) = \frac{3}{8}$

The probability of getting 1 head is

$P(\text{getting Rs } 100) = \frac{3}{8}$

Probability of getting 3 tails is

$P(\text{losing Rs } 1500) = \frac{1}{8}$

The probability distribution table is

x_i	-1500	100	300	500
$P(x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$$\text{Gain} = \text{Mean } \mu = \sum_{i=1}^{n} x_i P_i$$

$$= -1500\left(\frac{1}{8}\right) + 100\left(\frac{3}{8}\right) + 300\left(\frac{3}{8}\right) + 500\left(\frac{1}{8}\right)$$

$$= \frac{1}{8}\left[-1500 + 300 + 900 + 500\right]$$

$$= \frac{\cancel{200}\,\cancel{50}\,25}{\cancel{8}\,\cancel{2}\,1}$$

$$= 25$$

∴ The player gains Rs. 25

11) A fair die is tossed. Let the random Variable x denote twice sum of same number appearing on the die.

a) Write the probability distribution of x

b) Find Mean & Variance

Sol: Given a die is tossed

Total no. of cases = 6

Let x denote twice the sum of same number appearing on the die

a)
The probability distribution of x

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

b)
$$\text{Mean} = \sum_{i=1}^{n} x_i P_i$$

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

$$= \frac{1}{6}(42)$$

$$\mu = 7$$

$$\text{Variance } \sigma^2 = \sum_{i=1}^{n} x_i^2 P_i - \mu^2$$

$$= 4\left(\frac{1}{6}\right) + 16\left(\frac{1}{6}\right) + 36\left(\frac{1}{6}\right) + 64\left(\frac{1}{6}\right) + 100\left(\frac{1}{6}\right) + 144\left(\frac{1}{6}\right) - (7)^2$$

$$= \frac{1}{6}(364) - 7^2$$

$$= \frac{\overset{182}{\cancel{364}}}{\cancel{6}3} - 49$$

$$= \frac{182}{3} - 49$$

$$= 60.66 - 49$$

$$= 11.66$$

12) A random sample with replacement of size 2 is taken from $S = \{1,2,3\}$. Let the random variable X denote the sum of two numbers taken. Find

a) Probability Distribution

b) Mean

c) Variance

Sol: Given, a sample of size 2 is taken from $S = [1,2,3]$ with replacement

i.e $S = \{(1,1)\ (1,2)\ (1,3)\ (2,1)\ (2,2)\ (2,3)\ (3,1)\ (3,2)\ (3,3)\}$

Total no. of cases = 9

a) Probability Distribution

x_i	2	3	4	5	6
$P(x_i)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

b) Mean $\mu = \sum_i x_i P_i$

$$= 2\left(\frac{1}{9}\right) + 3\left(\frac{2}{9}\right) + 4\left(\frac{3}{9}\right) + 5\left(\frac{2}{9}\right) + 6\left(\frac{1}{9}\right)$$

$$= \frac{2+6+12+10+6}{9}$$

$= \frac{36}{9}$

$\mu = 4$

c) Variance $\sigma^2 = \sum x_i^2 p_i - \mu^2$

$= 4\left(\frac{1}{9}\right) + 9\left(\frac{2}{9}\right) + 16\left(\frac{3}{9}\right) + 25\left(\frac{2}{9}\right) + 36\left(\frac{1}{9}\right) - (4)^2$

$= \frac{4+18+48+50+36}{9} - 16$

$= \frac{156}{9} - 16$

$= \frac{\cancel{12}\,4}{\cancel{9}\,3}$

$= 1.33$

H.W

13) From a lot of 10 items containing 3 defective, a sample of 4 items are drawn at random. Let the random variable X denote no. of defective items. Find the probability distribution, mean & variance.

14) A player tosses two fair coins. He wins Rs 100 if one head appear, Rs 200 if 2 heads appear. On the other hand he loses ₹ 500 if no head appears. Determine the expected value of the game and is the game is favorable to the player.

13)

Sol: Let X denote the no. of defective items

Given, total number of items = 10

No. of defective items = 3

No. of good items = 10 − 3 = 7

Let X take the values 0, 1, 2, 3

$P(X=0) = P$(no defective item)

$$= \frac{{}^{7}C_{4}}{{}^{10}C_{4}}$$

$$= \frac{\cancel{7}\times\cancel{6}\times\cancel{5}\times\cancel{4}}{\cancel{10}\times\cancel{9}\times\cancel{8}\times\cancel{7}}$$

$$= \frac{1}{6}$$

$P(X=1) = P$(1 defective & 3 good items)

$$= \frac{{}^{3}C_{1}\times{}^{7}C_{3}}{{}^{10}C_{4}}$$

$$= \frac{\cancel{3}\times\cancel{7}\times\cancel{6}\times 5\times\cancel{4\times3\times2\times1}}{\cancel{10}\times\cancel{9}\times\cancel{8}\times\cancel{7}\times\cancel{3\times2\times1}}$$

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

$P(X=2) = P$(2 defective & 2 good items)

$$= \frac{{}^{3}C_{2}\times{}^{7}C_{2}}{{}^{10}C_{4}}$$

$$= \frac{\cancel{3}^{1} \times \cancel{7} \times \cancel{6}^{2\,1} \times \cancel{4} \times 3 \times \cancel{2} \times 1}{\cancel{10}_{5} \times \cancel{9}_{3} \times \cancel{8}_{1} \times \cancel{7} \times 2}$$

$$= \frac{3}{10}$$

$P(X=3) = P$ (3 defective items)

$$= \frac{{}^{3}C_{3} \times {}^{7}C_{1}}{{}^{10}C_{4}}$$

$$= \frac{1 \times \cancel{7} \times \cancel{4} \times \cancel{3}^{1} \times \cancel{2} \times 1}{10 \times \cancel{9}_{3} \times \cancel{8} \times \cancel{7}}$$

$$= \frac{1}{30}$$

∴ Probability Distribution

x	0	1	2	3
P(x)	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{30}$

Expected number of defective items $= E(x) = \sum x_i P_i$

$$\text{Mean} = 0 + \frac{1 \times 5}{2 \times 5} + 2\left(\frac{3}{10}\right) + \cancel{3}^{1}\left(\frac{1}{\cancel{30}_{10}}\right)$$

$$= \frac{5+6+1}{10}$$

$$= \frac{12}{10}$$

$$\mu = 1.2$$

Variance $(\sigma^2) = \sum x_i^2 P_i - \mu^2$

$$= 1\left(\frac{1}{2}\right) + 4\left(\frac{3}{10}\right) + \overset{3}{\cancel{9}}\left(\frac{1}{\cancel{30}_{10}}\right) - (1.2)^2$$

$$= \frac{1\times5}{2\times5} + \frac{12}{10} + \frac{3}{10} - (1.44)$$

$$= \frac{2\cancel{0}}{1\cancel{0}} - 1.44$$

$$\sigma^2 = 0.56$$

4)

Sol: Given 2 coins are tossed

Total no. of cases $= 2^2$
$= 4$

$S = \{HH, HT, TH, TT\}$

- Let X be the expected value of the game

The probability of getting 1 heads is

P(getting Rs 100) $= \frac{2}{4}$

The probability of getting 2 heads is

P(getting Rs 200) $= \frac{1}{4}$

The probability of getting no heads is

P(losing Rs 500) $= \frac{1}{4}$

Range of X $= \{-500, 100, 200\}$

X	-500	100	200
P(x)	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

$$\text{Mean} = \sum x_i p_i$$

$$= -500\left(\frac{1}{4}\right) + 100\left(\frac{2}{4}\right) + 200\left(\frac{1}{4}\right)$$

$$= \frac{1}{4}(-500+400)$$

$$= \frac{-100}{4}$$

$$= -25 \text{ Rs}$$

$\therefore$ The game is not favourable to the player.

29/09/21

Continuous Random Variables: A random variable X which takes values continuously i.e which takes all possible values in a given interval is called Continuous Random Variable.

Probability Density Function: Consider the interval $\left[x-\frac{dx}{2}, x+\frac{dx}{2}\right]$ of length dx around a 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 the interval $\left[x-\frac{dx}{2}, x+\frac{dx}{2}\right]$

Symbolically, it can be expressed as

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

Then $f(x)$ is called Probability Density Function.

Properties:

1) $f(x) \geq 0 \quad \forall\, x \in R$

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

Probability Distribution Function: The distribution function of a continuous random variable is denoted by 'x' and is defined as

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

Mean of a Continuous Random Variable: Mean of a continuous random variable is denoted as $E(x)$ or μ and is defined as

$$\boxed{\mu = \int_{-\infty}^{\infty} x f(x)\,dx}$$

Variance: Variance is denoted by σ^2 and is defined as

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

1) If a random variable has the probability density function as

$$f(x) = \begin{cases} 2e^{-2x} & x>0 \\ 0 & x\leq 0 \end{cases}$$

Find the probability

a) between 1 and 3

b) greater than 0.5

Sol: Given $f(x) = \begin{cases} 2e^{-2x} & x>0 \\ 0 & x\leq 0 \end{cases}$

a) $P(1 \leq x \leq 3)$

$$= \int_1^3 2e^{-2x}\,dx$$

$$= 2\int_1^3 e^{-2x}\,dx$$

$$= \cancel{2}\left[\frac{e^{-2x}}{-\cancel{2}}\right]_1^3$$

$$= -(e^{-6} - e^{-2})$$

$$= e^{-2} - e^{-6}$$

b) $P(x \geq 0.5)$

$$= \int_{0.5}^{\infty} 2e^{-2x}\,dx$$

$$= 2\int_{0.5}^{\infty} e^{-2x}\,dx$$

$$= \cancel{2}\left[\frac{e^{-2x}}{-\cancel{2}}\right]_0^{\infty}$$

$$= -\left(e^{-2x}\right)_{0.5}^{\infty}$$

$$= -\left[e^{-\infty} - e^{-2(0.5)}\right]$$

$$= e^{-1} - \frac{1}{e^{\infty}}$$

$$= e^{-1} - \frac{1}{\infty}$$

$$= e^{-1}$$

$$= \frac{1}{e}$$

NOTE

1) $e^{\infty} = \infty$
2) $e^{-\infty} = 0$
3) $e^{0} = 1$
4) $e^{-0} = 1$

2) If the probability density of a random variable is given by

$$f(x) = \begin{cases} K(1-x^2) & \text{for } 0<x<1 \\ 0 & \text{otherwise} \end{cases}$$

Find a) the value of K

b) Probability between 0.1 and 0.2

c) Greater than 0.5

d) Mean e) Variance

Sol: a) To find K:

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

$$\int_{0}^{1} f(x)dx = 1 \Rightarrow \int_{0}^{1} K(1-x^2)dx = 1$$

$$\Rightarrow K\left[x - \frac{x^3}{3}\right]_0^1 = 1$$

$$\Rightarrow K\left[\left(1-\frac{1}{3}\right) - (0-0)\right] = 1$$

$$\Rightarrow \frac{2K}{3} = 1 \quad \Rightarrow K = \frac{3}{2}$$

b) $$P(0.1 \le X \le 0.2) = \int_{0.1}^{0.2} \frac{3}{2}(1-x^2)\,dx$$

$$= \frac{3}{2}\left[x - \frac{x^3}{3}\right]_{0.1}^{0.2}$$

$$= \frac{3}{2}\left[(0.2-0.1) - \frac{1}{3}\left((0.2)^3-(0.1)^3\right)\right]$$

$$= \frac{3}{2}\left[0.1 - \frac{1}{3}\left(\frac{7}{1000}\right)\right]$$

$$= \frac{3}{2}\left[0.1\right] - \frac{\cancel{3}}{2}\left(\frac{1}{\cancel{3}}\right)\frac{7}{1000}$$ 00 0003·5, $\times 10^{-4}$

$$= 0.15 - 0.0035$$

$$= 0.1465$$

c) $$P(X \ge 0.5) = \int_{0.5}^{1} \frac{3}{2}(1-x^2)\,dx$$

$$= \frac{3}{2}\left[x - \frac{x^3}{3}\right]_{0.5}^{1}$$

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

$$= \frac{3}{2}\left[\frac{1}{2} - \frac{1}{3}\cdot\frac{7}{8}\right]$$

$$= \frac{3}{2}\left[0.2083\right]$$

$$= 0.3125$$

d) Mean $= \mu = \int_{-\infty}^{\infty} x\, f(x)\,dx$

$= \int_0^1 x\, f(x)\,dx$

$= \int_0^1 x\, \frac{3}{2}(1-x^2)\,dx$

$= \frac{3}{2}\int_0^1 (x - x^3)\,dx$

$= \frac{3}{2}\left[\frac{x^2}{2} - \frac{x^4}{4}\right]_0^1$

$= \frac{3}{2}\left[\left(\frac{1}{2}-0\right) - \left(\frac{1}{4}-0\right)\right]$

$= \frac{3}{2}\left(\frac{1}{4}\right)$

$= \frac{3}{8}$

$= 0.375$

e) Variance $\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$

$= \int_0^1 x^2\, \frac{3}{2}(1-x^2)\,dx - (0.375)^2$

$= \frac{3}{2}\int_0^1 (x^2 - x^4)\,dx - (0.375)^2$

$= \frac{3}{2}\left[\frac{x^3}{3} - \frac{x^5}{5}\right]_0^1 - (0.375)^2$

$= \frac{3}{2}\left(\frac{1}{3} - \frac{1}{5}\right) - (0.375)^2$

$= \frac{\cancel{3}}{\cancel{2}}\left(\frac{\cancel{2}}{\cancel{15}}\right) - 0.140625$

$= 0.059375 = 0.06$

3) If a random variable has the probability density function

$$f(x) = \begin{cases} K(x^2-1) & -1 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

Find the value of K and $P\left(\frac{1}{2} < x < \frac{5}{2}\right)$

<u>Sol:</u> <u>To find K</u>

$$\int_{-1}^{3} f(x)\,dx = 1 \quad \Rightarrow \int_{-1}^{3} K(x^2-1)\,dx = 1$$

$$\Rightarrow K\left(\frac{x^3}{3} - x\right)_{-1}^{3} = 1$$

$$\Rightarrow K\left[\frac{1}{3}\left(3^3-(-1)^3\right) - \left(3-(-1)\right)\right] = 1$$

$$\Rightarrow K\left[\frac{28}{3} - 4\right] = 1 \Rightarrow K\left[\frac{16}{3}\right] = 1$$

$$\Rightarrow K = \frac{3}{16}$$

$$P\left(\frac{1}{2} < x < \frac{5}{2}\right) = \int_{\frac{1}{2}}^{\frac{5}{2}} K(x^2-1)\,dx$$

$$= \frac{3}{16}\left[\frac{x^3}{3} - x\right]_{\frac{1}{2}}^{\frac{5}{2}}$$

$$= \frac{\cancel{3}}{16} \times \frac{1}{\cancel{3}}\left[\left(\frac{5}{2}\right)^3 - \left(\frac{1}{2}\right)^3\right] - \frac{3}{16}\left[\frac{5}{2} - \frac{1}{2}\right]$$

$$= \frac{1}{16}\left(\frac{\overset{31}{\cancel{124}}}{\underset{2}{\cancel{8}}}\right) - \frac{3}{16}(2)$$

$= \frac{31}{32} - \frac{6}{16}$

$= \frac{19}{32}$

$= 0.59$

Even Function: A function $y = f(x)$ is an even function if $\boxed{f(-x) = f(x)}$

Eg: Let $f(x) = x^2$

$f(-x) = (-x)^2$

$= x^2$

$= f(x)$

Odd Function: A function $y=f(x)$ is an odd function if $\boxed{f(-x) = -f(x)}$

Eg: Let $f(x) = x$

$f(-x) = -x$

$= -f(x)$

NOTE:

1) If $f(x)$ is an even function in the interval $-\infty < x < \infty$, in this case convert the integral as $2\int_0^{\infty} f(x)dx$

2) If $f(x)$ is an odd function in the interval $-\infty < x < \infty$ then the integral value is always zero

****

(4) The probability density function $f(x)$ of a continuous random variable is given by

$$f(x) = c e^{-|x|} \quad -\infty < x < \infty$$

Show that $c = \frac{1}{2}$ and find mean & variance of the distribution also find the probability between 0 & 4.

<u>Sol:</u> Given $f(x) = c e^{-|x|}$

<u>To find c:</u>

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \Rightarrow \int_{-\infty}^{\infty} c e^{-|x|}\,dx = 1$$

$$\Rightarrow c \int_{-\infty}^{\infty} e^{-|x|}\,dx = 1$$

Here $e^{-|x|}$ is an even function

$$\Rightarrow 2c \int_{0}^{\infty} e^{-|x|}\,dx = 1$$

$$\Rightarrow c\left[2 \int_{0}^{\infty} e^{-x}\,dx\right] = 1$$

$$\Rightarrow 2c \left(\frac{e^{-x}}{-1}\right)_0^{\infty} = 1$$

$$\Rightarrow -2c\left(e^{-\infty} - e^{-0}\right) = 1$$

$$\Rightarrow -2c(0-1) = 1$$

$\Rightarrow 2c = 1$

$\Rightarrow c = \frac{1}{2}$

Mean $\mu = \int_{-\infty}^{\infty} x\, f(x)\, dx$

$= \int_{-\infty}^{\infty} x\, c e^{-|x|}\, dx$

$= \frac{1}{2} \int_{-\infty}^{\infty} x\, e^{-|x|}\, dx$

$= \frac{1}{2} \int_{-\infty}^{\infty} x e^{-|x|}\, dx$

$x e^{-|x|}$ is an odd function

$\mu = 0$

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

$= \int_{-\infty}^{\infty} x^2 c\, e^{-|x|}\, dx - (0)^2$

$= \frac{1}{2} \int_{-\infty}^{\infty} x^2 e^{-|x|}\, dx$

$x^2 e^{-|x|}$ is an even function

$= \frac{1}{\cancel{2}} \times \cancel{2} \int_{0}^{\infty} x^2 e^{-|x|}\, dx$

$= \int_{0}^{\infty} x^2 e^{-x}\, dx$

$$\boxed{\begin{aligned}\int e^{x} f(x)\,dx &= e^{x}\left[f(x)-f'(x)+f''(x)-\cdots\right]\\ \int e^{-x} f(x)\,dx &= -e^{-x}\left[f(x)+f'(x)+f''(x)+\cdots\right]\end{aligned}}$$

$$= -e^{-x}\left[x^2+2x+2\right]_0^{\infty}$$

$$\cancel{= -\left(e^{-\infty}-e^{-0}\right)\left[(\infty-\infty+2)-(0-0-2)\right]}$$

$$\cancel{= 1(}$$

$$= \left(-x^2e^{-x} - 2xe^{-x} - 2e^{-x}\right)_0^{\infty}$$

$$= (\infty-\infty-\infty)-(0-0-2)$$

$$= 2$$

$$P(0\le x\le 4) = \int_0^4 c\,e^{-|x|}\,dx$$

$$= \frac{1}{2}\int_0^4 e^{-x}\,dx$$

$$= \frac{1}{2}\left(\frac{e^{-x}}{-1}\right)_0^4$$

$$= \frac{-1}{2}\left(e^{-4}-e^{-0}\right)$$

$$= \frac{-1}{2}\left[\frac{1}{e^4}-1\right]$$

$$= \frac{1}{2}\left[1-\frac{1}{e^4}\right] = \frac{1}{2}\left[1-\frac{1}{54.59}\right] = 0.4908$$

5) A continuous random variable has the probability density function

$$f(x) = \begin{cases} Kxe^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

a) Find K

b) Mean

c) Variance

Sol:- To find K

$$\int_{-\infty}^{\infty} f(x)dx = 1 \Rightarrow \int_0^{\infty} Kxe^{-\lambda x}dx = 1$$

$$\Rightarrow K\int_0^{\infty} xe^{-\lambda x}dx = 1$$

$$\Rightarrow K\left[x\left(\frac{e^{-\lambda x}}{-\lambda}\right) - 1\left(\frac{e^{-\lambda x}}{\lambda^2}\right)\right]_0^{\infty} = 1$$

$$\Rightarrow K\left[\frac{-xe^{-\lambda x}}{\lambda} - \frac{e^{-\lambda x}}{\lambda^2}\right]_0^{\infty} = 1$$

$$\Rightarrow K\left[(0-0) - \left(0-\frac{1}{\lambda^2}\right)\right] = 1$$

$$\Rightarrow K = \lambda^2$$

Mean $\mu = K\int_0^{\infty} x\,f(x)\,dx$

$$= \lambda^2\int_0^{\infty} x\,x\,e^{-\lambda x}dx$$

$$= \lambda^2\int_0^{\infty} x^2 e^{-\lambda x}dx$$

$$= \lambda^2 \left[x^2 \left(\frac{e^{-\lambda x}}{-\lambda} \right) - 2x \left(\frac{e^{-\lambda x}}{\lambda^2} \right) + 2 \left(\frac{e^{-\lambda x}}{-\lambda^3} \right) \right]_0^\infty$$

$$= \lambda^2 \left[-\frac{x^2 e^{-\lambda x}}{\lambda} - \frac{2x e^{-\lambda x}}{\lambda^2} - \frac{2 e^{-\lambda x}}{\lambda^3} \right]_0^\infty$$

$$= \lambda^2 \left[(\infty - \cancel{\infty}\,\infty) - \left(0 - 0 - \frac{2}{\lambda^3} \right) \right]$$

$$= \frac{2\lambda^2}{\lambda^3}$$

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

Variance $\sigma^2 = \int_0^\infty x^2 f(x) dx - \mu^2$

$$= \int_0^\infty x^2 K x e^{-\lambda x} dx - \left(\frac{2}{\lambda} \right)^2$$

$$= K \int_0^\infty x^3 e^{-\lambda x} dx - \left(\frac{2}{\lambda} \right)^2$$

$$= \lambda^2 \left[x^3 \left(\frac{e^{-\lambda x}}{-\lambda} \right) - 3x^2 \left(\frac{e^{-\lambda x}}{\lambda^2} \right) + 6x \left(\frac{e^{-\lambda x}}{-\lambda^3} \right) - 6 \left(\frac{e^{-\lambda x}}{\lambda^4} \right) \right]_0^\infty - \frac{4}{\lambda^2}$$

$$= \lambda^2 \left[0 - \left(0 - 0 + 0 - \frac{6}{\lambda^4} \right) \right] - \frac{4}{\lambda^2}$$

$$= \frac{6\lambda^2}{\lambda^4} - \frac{4}{\lambda^2}$$

$$= \frac{6}{\lambda^2} - \frac{4}{\lambda^2}$$

$$\sigma^2 = \frac{2}{\lambda^2}$$

H.W

(6) For the continuous random variable X whose probability density function is given by

$$f(x) = \begin{cases} c(2-x) & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

a) Find c b) mean c) Variance

H.W

(7) For the continuous probability function

$f(x) = kx^2 e^{-x}$ when $x \geq 0$. Find a) k b) Mean

c) Variance

**

(8) If X is a continuous random variable and $Y = aX + b$

Prove that $E(Y) = aE(X) + b$

Sol: Given X is a continuous random variable

By definition of expectation,

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

$$E(Y) = E(aX+b) = \int_{-\infty}^{\infty} (ax+b) f(x)\,dx$$

$$= \int_{-\infty}^{\infty} ax f(x)\,dx + \int_{-\infty}^{\infty} b f(x)\,dx$$

$$= a\int_{-\infty}^{\infty} x f(x)\,dx + b\int_{-\infty}^{\infty} f(x)\,dx$$

$$= a E(x) + b(1)$$

$$E(ax+b) = a E(x) + b$$

*****

9) If X is a continuous random Variable and K is any constant then prove that

a) $Var(X+K) = Var(X)$

b) $Var(Kx) = K^2 Var(X)$

Sol: PROOF - a) By definition of Variance, we have

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

$$= \int_{-\infty}^{\infty} x^2 f(x)dx - \left[\int_{-\infty}^{\infty} x f(x)dx\right]^2$$

$$Var(X+K) = \int_{-\infty}^{\infty} (x+K)^2 f(x)dx - \left[\int_{-\infty}^{\infty} (x+K) f(x)dx\right]^2$$

$$= \int_{-\infty}^{\infty} x^2 f(x)dx + \int_{-\infty}^{\infty} 2Kx f(x)dx + \int_{-\infty}^{\infty} K^2 f(x)dx - \left[\int_{-\infty}^{\infty} x f(x)dx + \int_{-\infty}^{\infty} K f(x)dx\right]^2$$

$$= E(x^2) + 2KE(x) + K^2(1) - \left[E(x) + K\right]^2$$

$$= E(x^2) + \cancel{2kE(x)} + \cancel{k^2} - \left[\left(E(x)\right)^2 + \cancel{k^2} + \cancel{2KE(x)}\right]$$

$$= E(x^2) - \left[E(x)\right]^2$$

$$\therefore Var(X+K) = Var(X)$$

b) $\text{var}(Kx) = \int_{-\infty}^{\infty} (Kx)^2 f(x)\, dx - \left[\int_{-\infty}^{\infty} Kx\, f(x) dx\right]^2$

$= K^2 \int_{-\infty}^{\infty} x^2 f(x) dx - \left[K \int_{-\infty}^{\infty} x f(x) dx\right]^2$

$= K^2 E(x^2) - \left[K E(x)\right]^2$

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

$= K^2 \left[E(x^2) - [E(x)]^2\right]$

$= K^2 \text{var}(x)$

⑥

Sol: a) To find $c \rightarrow$

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

$\Rightarrow \int_{0}^{2} c(2-x) dx = 1$

$\Rightarrow c\left[2x - \frac{x^2}{2}\right]_0^2 = 1$

$\Rightarrow c\left[2(2-0) - \frac{1}{2}(2^2 - 0^2)\right] = 1$

$\Rightarrow c[4-2] = 1 \Rightarrow 2c = 1$

$\Rightarrow c = \frac{1}{2}$

b) Mean $= \int_{-\infty}^{\infty} x f(x)\, dx$

$= \int_{0}^{2} x \left(\frac{1}{2}\right)(2-x)\, dx$

$= \frac{1}{2} \int_{0}^{2} (2x - x^2)\, dx$

$= \frac{1}{2}\left[\cancel{2}\left(\frac{x^2}{\cancel{2}}\right) - \frac{x^3}{3}\right]_0^2$

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

$= \frac{1}{2}\left[4 - \frac{8}{3}\right]$

$= \frac{1}{\cancel{2}} \times \frac{\cancel{4}^{2}}{3}$

$= \frac{2}{3}$

c) Variance $\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2$

$= \int_{0}^{2} \left(\frac{1}{2}\right) x^2 (2-x)\, dx - \left(\frac{2}{3}\right)^2$

$= \frac{1}{2} \int_{0}^{2} (2x^2 - x^3)\, dx - \frac{4}{9}$

$= \frac{1}{2}\left[\frac{2x^3}{3} - \frac{x^4}{4}\right]_0^2 - \frac{4}{9}$

$= \frac{1}{2}\left[\frac{2}{3}(2^3 - 0^3) - \frac{1}{4}(2^4)\right] - \frac{4}{9}$

$= \frac{1}{2}\left[\frac{16}{3} - 4\right] - \frac{4}{9}$

$$= \frac{1}{2}\left[\frac{4}{3}\right] - \frac{4}{9}$$

$$= \frac{4}{6} - \frac{4}{9}$$

$$= \frac{\cancel{12}\,2}{\cancel{6}_1 \times 9}$$

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

⑦

Sol: a) <u>To find K</u> →

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \Rightarrow \int_0^{\infty} Kx^2 e^{-x}\,dx = 1$$

$$\Rightarrow K\int_0^{\infty} x^2 e^{-x}\,dx = 1$$

$$\Rightarrow K\left[(-e^{-x})(x^2+2x+2)\right]_0^{\infty} = 1$$

$$\Rightarrow K\left[(\infty)(0) - (-1)(0+0+2)\right] = 1$$

$$\Rightarrow K(2) = 1 \quad \Rightarrow K = \frac{1}{2}$$

b) Mean $= \int_{-\infty}^{\infty} x f(x)\,dx$

$$= \int_0^{\infty} x\; Kx^2 e^{-x}\,dx$$

$$= K\int_0^{\infty} x^3 e^{-x}\,dx$$

$$= \frac{1}{2}\left[(-e^{-x})(x^3+3x^2+6x+6)\right]_0^{\infty}$$

$$= \frac{1}{2}\left[0(\infty) - (-1)(0+6)\right]$$

$= \frac{1}{2}(6)$

$= 3$

c) Variance $\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$

$= \int_{0}^{\infty} x^2 K x^2 e^{-x}\,dx - (3)^2$

$= \frac{1}{2} \int_{0}^{\infty} x^4 e^{-x}\,dx - 9$

$= \frac{1}{2}\left[-e^{-x}(x^4 + 4x^3 + 12x^2 + 24x + 24)\right]_0^{\infty} - 9$

$= \frac{1}{2}\left[(0) - (-1)(0+0+0+0+24)\right] - 9$

$= 12 - 9$

$= 3$

<u>Exhaustive Events</u>: A set of events are called exhaustive if atleast one of them necessarily occurs whenever the experiment is performed.

* If A starts the game

Probability of A's winning $= \frac{p}{1-q^2}$

Probability of B's winning $= \frac{pq}{1-q^2}$

04/10/21

UNIT-II

Probability Distributions

* We have already learnt frequency distributions which are based on the actual observations. In this chapter, we shall discuss theoretical distributions in which the variates are distributed as per some law which can be expressed mathematically.

* There are 2 types of probability distributions:

a) Discrete Probability Distribution

i) Binomial Distribution

ii) Poisson Distribution

b) Continuous Probability Distribution

Normal Distribution

Binomial Distribution: It was discovered by James Bernoulli in the year 1700 and it is a discrete Probability distribution.

* The probability of the no. of successes so obtained is called Binomial Probability Distribution.

Definition: A random variable X has a binomial distribution if it assumes only non-negative values & its probability density function is given by

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

Here n = no. of trials

p = Probability of success

q = Probability of failure

x = what to be found

Eg: 1) The no. of defective bolts in a box containing "n" bolts

2) The no. of machines lying idle in a factory having "n" machines

Conditions:

1) Trials are repeated under identical conditions for a fixed no. of times say "n" times.

2) There are only 2 possibilities i.e 2 possible outcomes

Eg: Success or Failure for each trial

3) The probability of success in each trial remains constant and doesnot change from trial to trial

4) The trials are independent i.e the probability of an event in any trial is not affected by the results

of any other trial.

NOTE: By Binomial Theorem,

① $\boxed{(q+p)^n = q^n + {}^nc_1 p q^{n-1} + {}^nc_2 q^{n-2} p^2 + \cdots + p^n}$

② The probability of success and failure is always equal to 1 i.e $\boxed{p+q=1}$

$$\Rightarrow q = 1-p$$

**
Mean of the Binomial Distribution: The mean of the binomial distribution is

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

Proof: By definition of Binomial Distribution, we have

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

By definition of mean,

$$\mu = \sum_{x=0}^{n} x . P(x)$$

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

$$= 0 + 1.{}^nc_1 p^1 q^{n-1} + 2.{}^nc_2 p^2 q^{n-2} + 3\,{}^nc_3 p^3 q^{n-3} + \cdots + n\left[{}^nc_n p^n q^{n-n}\right.$$

$$= npq^{n-1} + 2\frac{n(n-1)}{2!} p^2 q^{n-2} + 3\frac{n(n-1)(n-2)}{3!} p^3 q^{n-3} + \cdots + n(1)p^n($$

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

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

$$= np\left[q+p\right]^{n-1} \quad (\because \text{By Binomial Distribution})$$

$$= np(1)^{n-1}$$

$$\therefore \mu = np$$

★★★

Variance of Binomial Distribution: The variance of Binomial Distribution is

$$\boxed{\boxed{\sigma^2 = npq}}$$

Proof: By definition of Binomial Distribution, we have

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

By definition of Variance

$$\sigma^2 = \sum_{x=0}^{n} x^2 P(x) - \mu^2$$

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

$$= \sum_{x=0}^{n} \left[x(x-1) + x\right] {}^nC_x \, p^x q^{n-x} - \mu^2$$

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

$$= \left[0 + 0 + 2.1\,{}^nC_2\, p^2 q^{n-2} + 3.2.\,{}^nC_3 . p^3 q^{n-3} + \dots n(n-1)\,{}^nC_n\, p^n q^{n-n}\right] + \mu - \mu^2$$

$$= \left[\cancel{2}\,\frac{n(n-1)}{\cancel{2!}} p^2 q^{n-2} + \cancel{6}\,\frac{n(n-1)(n-2)}{\cancel{3!}} p^3 q^{n-3} + \dots + n(n-1)p^n\right] + \mu - \mu^2$$

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

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

$$= n(n-1)p^2\left[q+p\right]^{n-2} + \mu - \mu^2 \qquad (\because \text{By Binomial Distribution})$$

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

$$= (n^2 - n)p^2 + np - n^2p^2$$

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

$$= np(1-p)$$

$$\therefore \sigma^2 = npq$$

① A fair coin is tossed 6 times. Find the probability of getting 4 heads.

Sol: Given $n=6$

p = Probability of success

p = Probability of getting head $= \frac{1}{2}$

q = Probability of getting tail $= \frac{1}{2}$

$x = 4$

By Binomial Distribution, $P(x) = {}^{n}C_{x}\, p^{x} q^{n-x}$

$$= {}^{6}C_{4} \left(\frac{1}{2}\right)^{4} \left(\frac{1}{2}\right)^{6-4}$$

$$= \frac{\overset{3}{\cancel{6}}\times 5}{\underset{1}{\cancel{2}}\times 1} \left(\frac{1}{2}\right)^{6}$$

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

$$= \frac{15}{64}$$

$$= 0.234$$

② Determine the probability of getting the sum 6 exactly 3 times in 7 throws with a pair of fair die.

Sol: Total no. of cases $= 6^{2}$

$= 36$

$n = 7$

p = Probability of getting sum 6

$= \frac{5}{36}$

$q = 1 - p$

$= 1 - \frac{5}{36}$

$q = \frac{31}{36}$

$x = 3$

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

$= {}^7C_3 \left(\frac{5}{36}\right)^3 \left(\frac{31}{36}\right)^{7-3}$

$= \frac{7 \times 6 \times 5}{3 \times 2 \times 1} \times \frac{5^3}{36^3} \times \frac{31^4}{36^4}$

$= \frac{35 \times 125 \times 923521}{36^7}$

$= 0.0516$

③ Determine the probability of getting sum as 9 exactly twice in 3 throws with a pair of fair die.

<u>Sol:</u> Total no. of cases $= 6^2$

$= 36$

$x = 2, n = 3$

P = Probability of getting sum 9

$= \frac{4}{36}$

$= \frac{1}{9}$

$$q = 1 - p$$
$$= 1 - \frac{1}{9}$$
$$= \frac{8}{9}$$

By Binomial Distribution, $P(x) = {}^{n}C_{x}\, p^{x} q^{n-x}$

$$= {}^{3}C_{2} \left(\frac{1}{9}\right)^{2} \left(\frac{8}{9}\right)^{3-2}$$

$$= \cancel{3}^{1} \times \frac{1}{81} \times \frac{8}{\cancel{9}\,3}$$

$$= \frac{8}{243}$$

$$= 0.0329$$

$\boxed{05/10/21}$

(4) 10 coins are thrown simultaneously. Find probability of getting atleast

a) 7 heads

b) 6 heads

c) 1 head

NOTE:

1) Exactly - $P(X=x)$

2) Atleast - $P(X \geq x)$

3) Atmost - $P(X \leq x)$

4) More than - $P(X > x)$

Sol: Given $n=10$

P = probability of success (getting head)

$p=\frac{1}{2}$ $\quad$ q = Probability of getting tail $=1-\frac{1}{2}$

$=\frac{1}{2}$

a) Atleast 7 heads

$$P(x\geq 7)= P(x=7)+P(x=8)+P(x=9)+P(x=10)$$

$$= {}^{10}C_7\left(\frac{1}{2}\right)^{7}\left(\frac{1}{2}\right)^{10-7}+{}^{10}C_8\left(\frac{1}{2}\right)^{8}\left(\frac{1}{2}\right)^{10-8}+{}^{10}C_9\left(\frac{1}{2}\right)^{9}\left(\frac{1}{2}\right)^{10-9}+{}^{10}C_{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^{0}$$

$$= {}^{10}C_3\left(\frac{1}{2}\right)^{10}+{}^{10}C_2\left(\frac{1}{2}\right)^{10}+\left(\frac{1}{2}\right)^{10}{}^{10}C_1+{}^{10}C_0\left(\frac{1}{2}\right)^{10}$$

$$=\left(\frac{1}{2}\right)^{10}\left[\frac{10\times\cancel{9}^{3}\times\cancel{8}^{4}}{\cancel{3}\times\cancel{2}\times1}+\frac{\cancel{10}^{5}\times9}{\cancel{2}\times1}+10+1\right]$$

$$=\frac{1}{2^{10}}\left[120+45+11\right]$$

$$=\frac{176}{2^{10}}$$

$$=0.1719$$

b) Atleast 6 heads

$$P(x\geq 6)=P(x=6)+P(x=7)+P(x=8)+P(x=9)+P(x=10)$$

$$= {}^{10}C_6\left(\frac{1}{2}\right)^{6}\left(\frac{1}{2}\right)^{10-6}+\frac{176}{2^{10}}$$

$$=\left[\frac{10\times\cancel{9}^{3}\times\cancel{8}\times7}{\cancel{4}\times\cancel{3}\times\cancel{2}\times1}+176\right]\frac{1}{2^{10}}$$

$$=\frac{(210+176)}{2^{10}} = 0.3769$$

c) Atleast 1 head

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

$$= 1 - {}^{10}C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{10-0}$$

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

$$= 0.9990$$

⑤ Two dice are thrown 5 times, find the probability of getting 7 as sum

a) Atleast once

b) Two times

c) $P(1<x<5)$

Sol: Given 2 dice are thrown

$\therefore$ No. of possible cases $= 6^2 = 36$

$n = 5$

P = Probability of getting sum as 7

$= \frac{6}{36}$

$P = \frac{1}{6}$

$$q = 1 - P$$

$$= 1 - \frac{1}{6}$$

$$= \frac{5}{6}$$

a) Atleast once

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

$= 1 - {}^5C_0 \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^{5-0}$

$= 1 - 1\left(\frac{5}{6}\right)^5$

$= 0.5981$

b) Two times

$P(x=2) = {}^5C_2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{5-2}$

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

$= \frac{1250}{7776}$

$= 0.1607$

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

$= {}^5C_2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{5-2} + {}^5C_3 \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^{5-3} + {}^5C_4 \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^{5-4}$

$= \frac{1}{6^5}\left[\frac{5 \times \cancel{4}^{2}}{\cancel{2}_{1} \times 1} \times 5^3 + \frac{5 \times \cancel{4}^{2}}{\cancel{2}_{1} \times 1} \times 5^2 + 5 \times 5\right]$

$= \frac{1}{6^5}\left[1250 + 250 + 25\right]$

$= \frac{1525}{6^5}$

$= 0.1961$

⑥ 20% of items produced from a factory are defective. Find the probability that in a sample of 5 chosen at random

a) None is defective

b) One is defective

c) $P(1<X<4)$

Sol: Given $n = 5$

P = Probability of getting defective item

$= \frac{\cancel{20}}{\cancel{100}_{5}}$

$= \frac{1}{5}$

$q = 1 - \frac{1}{5}$

$= \frac{4}{5}$

a) None is defective

$P(X=0) = {}^5C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^{5-0}$

$= 1 \times \left(\frac{4}{5}\right)^5$

$= 0.3276$

b) One is defective

$P(X=1) = {}^5C_1 \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^{5-1}$

$= \cancel{5} \times \frac{1}{\cancel{5}} \times \left(\frac{4}{5}\right)^4 = 0.4096$

c) $P(1<x<4) = P(X=2) + P(X=3) + \cancel{P(X=4)}$

$$= {}^5C_2\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^{5-2} + {}^5C_3\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^{5-3}$$

$$= \frac{5\times\cancel{4}^{\,2}}{\cancel{2}_{\,1}\times1}\times\frac{4^3}{5^5} + \frac{5\times\cancel{4}^{\,2}}{\cancel{2}_{\,1}\times1}\times\frac{4^2}{5^5}$$

$$= \frac{640+160}{5^5}$$

$$= 0.256$$

⑦ Assume that 50% of all engineering students are good in Mathematics. Determine the probabilities that among 18 engineering students

a) exactly 10

$= 0.1669$

b) Atleast 10

$$P(x \geq 10) = P(x=11) + P(x=12) + P(x=13) + P(x=14) + P(x=15) + P(x=16)$$
$$+ P(x=17) + P(x=18) + P(x=10)$$

$$= {}^{18}C_{11}\left(\tfrac{1}{2}\right)^{18} + {}^{18}C_{12}\left(\tfrac{1}{2}\right)^{18} + {}^{18}C_{13}\left(\tfrac{1}{2}\right)^{18} + {}^{18}C_{14}\left(\tfrac{1}{2}\right)^{18} + {}^{18}C_{15}\left(\tfrac{1}{2}\right)^{18} + {}^{18}C_{16}\left(\tfrac{1}{2}\right)^{18}$$
$$+ {}^{18}C_{17}\left(\tfrac{1}{2}\right)^{18} + {}^{18}C_{18}\left(\tfrac{1}{2}\right)^{18} + {}^{18}C_{10}\left(\tfrac{1}{2}\right)^{18}$$

$$= \left(\tfrac{1}{2}\right)^{18}\left[{}^{18}C_{10} + {}^{18}C_{11} + {}^{18}C_{12} + {}^{18}C_{13} + {}^{18}C_{14} + {}^{18}C_{15} + {}^{18}C_{16} + {}^{18}C_{17} + {}^{18}C_{18}\right]$$

$= 0.407$

c) Atmost 8

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

$$= \left[{}^{18}C_0 + {}^{18}C_1 + {}^{18}C_2 + {}^{18}C_3 + {}^{18}C_4 + {}^{18}C_5 + {}^{18}C_6 + {}^{18}C_7 + {}^{18}C_8\right]\left(\tfrac{1}{2}\right)^{18}$$

$$= \frac{106762}{2^{18}}$$

$= 0.407$

d) $P(2 \le x \le 9) = P(x=2) + P(x=3) + P(x=4) + P(x=5) + P(x=6) + P(x=7)$

$+ P(x=8) + P(x=9)$

$$= \left({}^{18}C_2 + {}^{18}C_3 + {}^{18}C_4 + {}^{18}C_5 + {}^{18}C_6 + {}^{18}C_7 + {}^{18}C_8 + {}^{18}C_9\right)\left(\frac{1}{2}\right)^{18} = \frac{155363}{2^{18}}$$

$= 0.5926$

06/10/21

8) If the probability of a defective bolt is $\frac{1}{8}$. Find

a) Mean

b) Variance of the distribution of the defective bolt of 640

Sol: $n = 640$

P = Probability of defective bolt

$p = \frac{1}{8}$ $\qquad$ $q = 1 - \frac{1}{8}$

$= \frac{7}{8}$

Mean $\mu = np$

$= 640\left(\frac{1}{8}\right)$

$= 80$

Variance $\sigma^2 = npq$

$= \overset{10}{\cancel{640}}\left(\frac{1}{\cancel{8}}\right)\left(\frac{7}{\cancel{8}}\right)$

$= 70$

(9) If the probability of a defective bolt is 0.2. Find

a) Mean b) Standard Deviation for the distribution

of bolts in a total of 400

<u>Sol:</u> Given $n = 400$

$$P = \frac{2}{10} = \frac{1}{5} \qquad q = 1 - \frac{1}{5} = \frac{4}{5}$$

Mean $\mu = np$

$$= 400\left(\frac{1}{5}\right)$$

$$= 80$$

Variance $\sigma^2 = npq$

$$= \cancel{400}^{\cancel{80}\,16} \times \frac{1}{\cancel{5}_1} \times \frac{4}{\cancel{5}_1}$$

$$= 64$$

$$S.D = \sqrt{\text{Variance}}$$

$$= \sqrt{64}$$

$$= 8$$

(10) The mean and variance of a binomial distribution are 4 and $\frac{4}{3}$ respectively. Find $P(X \geq 1)$

<u>Sol:</u> Given $\mu = np = 4$ — (1)

$$\sigma^2 = npq = \frac{4}{3} \quad \text{— (2)}$$

$$\frac{(1)}{(2)} \Rightarrow \frac{\cancel{np}}{\cancel{np}q} = \frac{\cancel{4}}{\cancel{4}} \times 3 \Rightarrow \frac{1}{q} = 3$$

$\Rightarrow q = \frac{1}{3}$ $\quad$ $P = 1 - q$ $\quad$ ① $\Rightarrow np = 4$

$= 1 - \frac{1}{3}$ $\quad$ $n\left(\frac{\cancel{2}^{1}}{3}\right) = \cancel{4}^{2}$

$= \frac{2}{3}$ $\quad$ $n = 6$

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

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

$$= 1 - \frac{1}{3^6}$$

$$= 0.998$$

⑪ Determine the Binomial Distribution for which mean is 4 and Variance is 3

<u>Sol:</u> $\mu = np = 4$ — ①

$\sigma^2 = npq = 3$ — ②

$$\frac{②}{①} \Rightarrow \frac{\cancel{n}pq}{\cancel{n}p} = \frac{3}{4} \Rightarrow q = \frac{3}{4}$$

$$P = 1 - q = 1 - \frac{3}{4} = \frac{1}{4}$$

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

***

⑫ 6 dice are thrown, 729 times. How many times do you expect atleast 3 dice to show 5 or 6

<u>Sol:</u> Given $n = 6$

P = Probability of getting 5 or 6

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

$P = \frac{1}{3}$ $\quad$ $q = 1 - \frac{1}{3} = \frac{2}{3}$

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

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

$$= 1 - \left[{}^6C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^{6-0} + {}^6C_1 \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^{6-1} + {}^6C_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^{6-2} \right]$$

$$= 1 - \left[\left(\frac{2}{3}\right)^6 + \frac{6 \times 2^5}{3^6} + \frac{15 \times 2^4}{3^6} \right]$$

$$= 1 - \frac{1}{3^6}\left[64 + 192 + 240 \right]$$

$$= 1 - \frac{496}{3^6}$$

$$= 0.3196$$

$$= \frac{233}{729}$$

$\therefore$ The expected no. of such cases in 729 times

is $729 \times \frac{233}{729} = 233$

③ Two dice are thrown 120 times. Find the average no. of times in which the number on the first die exceeds the number on the second die.

Sol: Given 2 dice are thrown i.e $n=2$

Total no. of cases $= 6^2 = 36$

P = Probability of getting a number which ~~exists~~ excee
second die

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

$= \frac{\cancel{15}\,5}{\cancel{36}\,12}$

$P = \frac{5}{12}$ $\qquad\qquad q = 1 - \frac{5}{12} = \frac{7}{12}$

Mean $\mu = np$ $\qquad\qquad n = 2$

$\cancel{\boxed{= \overset{10}{\cancel{120}} \times \frac{5}{\cancel{12}_1} \quad = 50}}$ $\qquad = 2 \times \frac{5}{12}$

$= \frac{5}{6}$

$= \frac{5}{\cancel{6}_1} \times \overset{20}{\cancel{120}}$

$= 100$

(14) Out of 800 families with 5 children each how many do 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 and girls

Sol: Given $n=5$

P = Prob. of boy $=\frac{1}{2}$

q = Prob. of girl $=\frac{1}{2}$

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

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

$= \frac{5}{16}$ → Only for one family

Hence the Probability of no. of families having 3 boys is

For 800 families, $\frac{5}{\cancel{16}_1} \times \cancel{800}^{50}$

$=250$

b) $P(x=5) = {}^5C_5\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^{5-5}$

$= 1\times\frac{1}{2^5}$

$= \frac{1}{32}$

For 800 families $= \frac{1}{32}\times 800$

$=25$

c) $P(x=2)+P(x=3) = {}^5C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{5-2} + {}^5C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{5-3}$

$= \cancel{2}\times\frac{5\times 4}{\cancel{2}\times 1}\left(\frac{1}{2^5}\right)$

$= \frac{\cancel{20}^{10}}{\cancel{32}_{16}}$

$= \frac{10}{16}$

For 800 families, $\frac{10}{\cancel{16}_1} \times \cancel{800}^{50}$

$= 500$

d) $P(x \geq 1) = 1 - P(x=0)$

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

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

$= \frac{31}{32}$

For 800 families, $\frac{31}{32} \times 800$

$= 775$

⑮ Fit a Binomial Distribution to the following frequency distribution

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

Sol: Given $n = 6$

$N =$ Total frequency

$= \sum f_i$

$= 13 + 25 + 52 + 58 + 32 + 16 + 4$

$= 200$

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i} = \frac{0(13) + 1(25) + 2(52) + 3(58) + 4(32) + 5(16) + 6(4)}{200}$$

$= \frac{535}{200}$

$= 2.675$

$np = 2.675 \Rightarrow p = \frac{2.675}{6}$

$\Rightarrow p = 0.445$

$q = 1 - p$

$= 1 - 0.445$

$q = 0.555$

x	f	$P(x) = {}^nC_x q^{n-x} p^x$	Expected Frequency $E = N \times P(x)$	Rounding off
0	18	$P(0) = {}^6C_0 (0.555)^6 (0.445)^0$ $= 0.029$	$= 200 \times P(0)$ $= 200 \times (0.029)$ $= 5.84$	6
1	25	$P(1) = {}^6C_1 (0.555)^5 (0.445)^1$ $= 0.14$	$= 200 \times P(1)$ $= 28.11$	28
2	52	$P(2) = {}^6C_2 (0.555)^4 (0.445)^2$ $= 0.28$	$= 200 \times P(2)$ $= 56.36$	56
3	58	$P(3) = {}^6C_3 (0.555)^3 (0.445)^3$ $= 0.30$	$= 200 \times P(3)$ $= 60.25$	60
4	32	$P(4) = {}^6C_4 (0.555)^2 (0.445)^4$ $= 0.18$	$= 200 \times P(4)$ $= 36.23$	36
5	16	$P(5) = {}^6C_5 (0.555)^1 (0.445)^5$ $= 0.058$	$= 200 \times P(5)$ $= 11.62$	12
6	4	$P(6) = {}^6C_6 (0.555)^0 (0.445)^6$ $= 7.765\ldots \times 10^{-3}$	$= 200 \times P(6)$ $= 1.55$	2

Expected Probability $= 6+28+56+60+36+12+2$
(Frequency)
$= 200$

$=$ Given Probability (Frequency)
(or)
Actual Probability (Frequency)

6) 4 coins are tossed 160 times. The no. of times head occurs is given below

x	0	1	2	3	4
No. of times	8	34	69	43	6

Fit a Binomial Distribution to this data on the hypothesis that the coins are unbiased

Sol: Given $n=4$

$\therefore$ Given coins are unbiased $\Rightarrow p=\frac{1}{2}$ $q=\frac{1}{2}$

x	f	$P(x) = {}^nC_x\, q^{n-x}\, p^x$	Expected Frequency $E = N \times P(x)$	Rounding Off
0	8	$P(0) = {}^4C_0 \left(\frac{1}{2}\right)^{4-0} \left(\frac{1}{2}\right)^0$ $= 0.0625$	$E = 160 \times P(0)$ $= 160 \times 0.0625$ $= 10$	10
1	34	$P(1) = {}^4C_1 \left(\frac{1}{2}\right)^{4-1} \left(\frac{1}{2}\right)^1$ $= 0.25$	$= 160 \times P(1)$ $= 160 \times 0.25$ $= 40$	40
2	69	$P(2) = {}^4C_2 \left(\frac{1}{2}\right)^{4-2} \left(\frac{1}{2}\right)^2$ $= 0.375$	$= 160 \times P(2)$ $= 160 \times 0.375$ $= 60$	60
3	43	$P(3) = {}^4C_3 \left(\frac{1}{2}\right)^{4-3} \left(\frac{1}{2}\right)^3$ $= 0.25$	$= 160 \times P(3)$ $= 160 \times 0.25$ $= 40$	40
4	6	$P(4) = {}^4C_4 \left(\frac{1}{2}\right)^{4-4} \left(\frac{1}{2}\right)^4$ $= 0.0625$	$= 160 \times P(4)$ $= 160 \times 0.0625$ $= 10$	10

Expected Frequency $= 10 + 40 + 60 + 40 + 10$

$= 160$

$=$ Actual Frequency

20/10/21

Poisson Distribution: Poisson Distribution is a distinct distribution. This distribution is applied when the events whose probability of occurence is very small but no. of trials are very large.

n is very large

p is very small

Definition: A random variable X is said to follow Poisson distribution if it assumes only non-negative values and its probability density function is given by

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

Eg: 1) The no. of telephone calls per min at a switch board

2) The no. of cars passing a certain point in one minute

3) The no. of printing mistakes per page in a large text book.

Conditions:

1) The no. of occurences should be a discrete variabl

2) The no. of trials 'n' should be large.

3) The no. of Printing mistakes per page in a large textbook.

3) The probability of success 'p' is very small

$np=\lambda$ is finite

Mean of Poisson Distribution: The mean of Poisson distribution is λ i.e $\lambda=np$

Proof → Let X is a discrete random Variable

By definition of poisson distribution, we have

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

By definition of mean $(\mu) = \sum_{x=0}^{n} x.P(x)$

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

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

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

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

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

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

$$\mu = \lambda$$

Variance of Poisson Distribution: The variance of Poisson distribution is λ

Proof $\rightarrow$ By definition of poisson distribution, we have

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

By definition of variance

$$Var(x) = \sum_{x=0}^{n} x^2 P(x) - \mu^2$$

$$= \sum_{x=0}^{n} x^2 \frac{e^{-\lambda}\lambda^{x}}{x!} - \lambda^2$$

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

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

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

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

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

$$= \lambda^2 e^{-\lambda}\left[\sum_{x=2}^{n} \frac{\lambda^{x-2}}{(x-2)!}\right] + \lambda - \lambda^2$$

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

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

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

$Var(x) = \lambda$

① If the probability that a person suffers a bad reaction from a certain injection is 0.001, determine the probability that out of 2000 members

a) Exactly 3

b) More than 2 persons

c) None

d) More than 1 person suffers a bad reaction

Sol: Given $p = 0.001$ $n = 2000$

By definition of poisson distribution,

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

We have $\lambda = np$

$= 2000(0.001)$

$= 2$

a) $P(x=3) = \dfrac{e^{-\lambda}\lambda^x}{x!}$

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

$= 0.1804$

b) $P(x>2) = 1 - P(x\leq 2)$

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

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

$$= 1-\left[0.135 + 0.270 + 0.270\right]$$

$$= 0.325$$

c) $P(X=0) = \dfrac{e^{-2}2^{0}}{0!}$

$$= 0.135$$

d) $P(x>1) = 1-P(x\leq 1)$

$$= 1-\left[P(x=0) + P(x=1)\right]$$

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

$$= 1-\left[0.135 + 0.270\right]$$

$$= 0.595$$

2) A manufacturer knows that condensers he makes contain an average 1% defectives. He packs them in boxes of 100. What is the probability that a box picked at random will contain 3 or more defective condensers.

Sol: Given $n=100$ $\quad P=1\%$

$$=\frac{1}{100}=0.01$$

$$\lambda=np$$
$$=100(0.01)$$
$$=1$$

$$P(X\geq 3)=1-P(X<3)$$
$$=1-[P(X=0)+P(X=1)+P(X=2)]$$
$$=1-\left[\frac{e^{-1}1^{0}}{0!}+\frac{e^{-1}1^{1}}{1!}+\frac{e^{-1}1^{2}}{2!}\right]$$
$$=1-[0.367+0.367+0.183]$$
$$=0.083$$

③ A hospital switch board receives an average of 4 emergency calls in 10 min, what is the probability that

a) There are atmost 2 emergency calls

b) There are exactly 3 emergency calls

Sol: $\lambda=4$

a) $P(X\leq 2)=P(X=0)+P(X=1)+P(X=2)$

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

$$=0.0183+0.073+0.146$$

$$=0.237$$

b) $P(X=3) = \dfrac{e^{-4}4^{3}}{3!}$

$= 0.195$

④ Average no. of accidents on any day on a natural highway is 1.8. Determine the probability that the no. of accidents are a) atleast one

b) atmost one

Sol: $\lambda = 1.8$

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

$= 1 - \left[\dfrac{e^{-1.8}(1.8)^{0}}{0!}\right]$

$= 1 - 0.165$

$= 0.835$

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

$= 0.165 + \dfrac{e^{-1.8}(1.8)^{1}}{1!}$

$= 0.165 + 0.297$

$= 0.462$

⑤ If the variance of a poisson distribution is 3. Find the probability that a) $x=0$

b) $0 < x \leq 3$

c) $1 \leq x \leq 4$

Sol: $\lambda = 3$

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

$= 0.049$

b) $P(0<x\le 3) = P(x=1) + P(x=2) + P(x=3)$

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

$= 0.149 + 0.224 + 0.224$

$= 0.597$

c) $P(1\le x\le 4) = P(x=1) + P(x=2) + P(x=3) + P(x=4)$

$= 0.149 + 0.224 + 0.224 + 0.168$

$= 0.765$

23/10/21

⑥ 2% of items of a factory are defective. The items are packed in boxes. What is the probability that there will be

a) 2 defective items

b) Atleast 3 defective items in a box of 100 items

Sol: $n = 100$ $\quad p = 2\%$

$= \frac{2}{100}$

$\lambda = np$

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

$\lambda = 2$

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

$= 2e^{-2}$

$= 0.2706$

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

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

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

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

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

$= 0.3233$

NOTE: The recurrence formula for Poisson Distribution is

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

2) If X is a Poisson Variable such that $P(x=0) = P(x=1)$, find $P(x=0)$ and using recurrence Formula find the probabilities at $x = 1, 2, 3, 4 \; \& \; 5$

Sol: Given $P(x=0)=P(x=1)$

$$\frac{e^{-\lambda}\lambda^0}{0!}=\frac{e^{-\lambda}\lambda^1}{1!}$$

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

$$\Rightarrow \lambda=1$$

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

$$=1.e^{-1}$$

$$=0.3678$$

$$\left[\because \text{By recurrence } P(x+1)=\frac{\lambda}{x+1}P(x)\right]$$

$$P(x=1)=P(x+1) \quad x=0$$

$$=\frac{1}{0+1}.0.3678$$

$$=0.3678$$

$$P(x=2)=\frac{1}{2}(0.3678) \qquad P(x=2)=\frac{1}{2}P(x=1)$$

$$=0.1839$$

$$P(x=3)=\frac{1}{3}(0.1839) \qquad P(x=3)=\frac{1}{3}P(x=2)$$

$$=0.0613$$

$$P(x=4)=\frac{1}{4}(0.0613) \qquad P(x=4)=\frac{1}{4}P(x=3)$$

$$=0.015325$$

$$x=5)=\frac{1}{5}(0.015325) \qquad P(x=5)=\frac{1}{5}P(x=4)$$

$$=0.003065 \quad =0.0030$$

3) Using recurrence formula, find the probabilities when $x = 0, 1, 2, 3, 4, 5$ if the mean of P.D is 3

Sol: $\mu = \lambda = 3$

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

$$= 0.0497$$

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

$$= 3(0.0497)$$

$$= 0.1491$$

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

$$= \frac{3}{2}(0.1491)$$

$$= 0.2236$$

$$P(x=3) = \frac{3}{3}(0.2236)$$

$$= 0.2236$$

$$P(x=4) = \frac{3}{4} P(x=3)$$

$$= \frac{3}{4}(0.2236)$$

$$= 0.1677$$

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

$$= \frac{3}{5}(0.1677)$$

$$= 0.100$$

⑨ If a P.D is such that $P(x=1) \cdot \frac{3}{2} = P(x=3)$

Find a) $P(x \geq 1)$

b) $P(x \leq 3)$

c) $P(2 \leq x \leq 5)$

Sol:

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

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

$$\Rightarrow \lambda^2 = \frac{3}{\cancel{2}_1} \times \cancel{6}^3$$

$$\Rightarrow \lambda^2 = 9 \Rightarrow \lambda = 3$$

ⓐ $P(x \geq 1) = 1 - P(x=0)$

$$= 1 - \left[\frac{e^{-3} 3^0}{0!}\right]$$

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

$$= 0.9502$$

ⓑ $P(x \leq 3) = P(x=0) + P(x=1) + P(x=2) + P(x=3)$

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

$$= e^{-3} + 3e^{-3} + \frac{9}{2}(e^{-3}) + \frac{9}{2}(e^{-3})$$

$$= 4e^{-3} + 9e^{-3}$$

$$= 13e^{-3}$$

$$= 0.6472$$

(c) $P(2 \le X \le 5) = P(X=2) + P(X=3) + P(X=4) + P(X=5)$

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

$$= \frac{9}{2} e^{-3} + \frac{9}{2} e^{-3} + \frac{27 \times 5}{8 \times 5} e^{-3} + \frac{81}{40} e^{-3}$$

$$= 9e^{-3} + \frac{216}{40} e^{-3}$$

$$= \frac{72}{5} e^{-3}$$

$$= 0.7169$$

(10) Fit a P.D for the following data and calculate the expected frequency

x	0	1	2	3	4
$F(x)$	109	65	22	3	1

Sol: $N = \sum f_i$ = Total frequency

$= 109 + 65 + 22 + 3 + 1$

$= 200$

Mean $\mu = \frac{\sum f_i x_i}{\sum f_i}$

$$= \frac{0 + 65 + 44 + 9 + 4}{200}$$

$$= \frac{122}{200}$$

$$= 0.61$$

$\lambda = 0.61$

x	$f(x)$	$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$	Expected Frequency $= N \times P(x)$	Rounding off
0	109	$P(x) = \frac{e^{-0.61}\lambda^0}{0!}$ $= 0.5433$	$= 200 \times 0.5433$ $= 108.6$	109
1	65	$P(x) = \frac{e^{-0.61}(0.61)}{1!}$ $= 0.3314$	$= 200 \times 0.3314$ $= 66.2$	66
2	22	$P(x) = \frac{e^{-0.61}(0.61)^2}{2!}$ $= 0.1010$	$= 200 \times 0.1010$ $= 20.2$	20
3	3	$P(x) = \frac{e^{-0.61}(0.61)^3}{3!}$ $= 0.0205$	$= 200 \times 0.0205$ $= 4.11$	4
4	1	$P(x) = \frac{e^{-0.61}(0.61)^4}{4!}$ $= 3.134 \times 10^{-3}$	$= 200 \times 3.134 \times 10^{-3}$ $= 0.62$	1

(11) The distribution of typing mistakes commited by a typist is given below.

Assuming the distribution to be poisson, find the expected frequencies.

x	0	1	2	3	4	5
$f(x)$	42	33	14	6	4	1

Sol:- $\sum f_i = 42 + 33 + 14 + 6 + 4 + 1$

$= 100$

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i} = \frac{0 + 33 + 28 + 18 + 16 + 5}{100}$$

$= 1$

x	$f(x)$	$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$	Expected Frequency $= 100 \times P(x)$	Rounding off.
0	42	$= \frac{e^{-1} 1^{0}}{0!}$ $= e^{-1} = 0.3678$	$= 100 \times 0.3678$ $= 36.78$	37
1	33	$= \frac{e^{-1} 1^{1}}{1!}$ $= 0.3678$	$= 100 \times 0.3678$ $= 36.78$	37
2	14	$= \frac{e^{-2}(1)^{2}}{2!}$ $= \cancel{0.0676}$ 0.1839	$= 100 \times \cancel{0.0676}$ 0.1839 $= \cancel{6.76}$ 18.39	18
3	6	$= \frac{e^{-1}(1)^{3}}{3!}$ $= 0.0061$	$= 100 \times 0.0061$ $= \cancel{0.6126}$ 6.131	6
4	4	$= \frac{e^{-1}(1)^{4}}{4!}$ $= 0.0153$	$= 100 \times 0.0153$ $= 1.53$	2
5	1	$= \frac{e^{-1}(1)^{5}}{5!}$ $= 3.06\ldots \times 10^{-3}$	$= 100 \times 3.06\ldots \times 10^{-3}$ $= 0.30$	0

25/10/21

Normal Distribution: The normal distribution is a continuous probability distribution. A continuous distribution is a distribution in which the variate can take all values within a given range.

Eg: The height of a person, the speed of the vehicle etc.

Definition: A random variable X is said to have a normal distribution if its density function is given by

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \begin{array}{c} -\infty < x < \infty \\ -\infty < \mu < \infty \\ \sigma > 0 \end{array}$$

Standard Normal Distribution: The normal distribution with mean $\mu = 0$ and standard Deviation $\sigma = 1$ is known as Standard Normal Distribution.

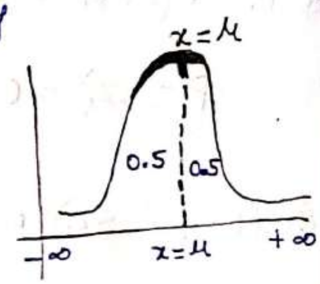
NOTE: The random variable that follows normal distribution is denoted by Z. If a variable X follows normal distribution with mean μ and S.D σ, then the variable Z is defined as

$$Z = \frac{x-\mu}{\sigma}$$

Characteristics of Normal Distribution:

1) The graph of Normal Distribution $y=f(x)$ in the xy plane is known as Normal Curve.

2) The curve is a bell shape curve and symmetrical with respect to mean i.e about the line $x=\mu$ and the two tails on the right and left side of the mean extends to infinity



3) Area under the normal curve represents the total population.

4) Mean, Median and Mode of the distribution coincide at $x=\mu$ as the distribution is symmetrical. So the normal curve is unimodel.

5) Linear combination of independent normal variables is also a normal variable.

Formulae:

① If both z_1 & z_2 are positive (or) negative then

$P(z_1 \leq z \leq z_2) = |A(z_2) - A(z_1)|$ where A is area

② If both Z_1 and Z_2 are of opposite sign then

$$P(Z_1 \le Z \le Z_2) = A(Z_2) + A(Z_1)$$

③ If $P(Z \geq Z_1)$ or $P(Z > Z_1)$ and $Z_1 > 0$ then

$$P(Z \geq Z_1) = 0.5 - A(Z_1)$$

④ If $P(Z \geq Z_1)$ or $P(Z > Z_1)$ and $Z_1 < 0$ then

$$P(Z \geq Z_1) = 0.5 + A(Z_1)$$

⑤ If $P(Z \le Z_1)$ or $P(Z < Z_1)$ and $Z_1 > 0$ then

$$P(Z \le Z_1) = 0.5 + A(Z_1)$$

⑥ If $P(Z \le Z_1)$ or $P(Z < Z_1)$ and $Z_1 < 0$ then

$$P(Z \le Z_1) = 0.5 - A(Z_1)$$

⑦ $A(-Z_1) = A(Z_1)$ [∵ The curve is Symmetric]

① For a normally distributed Variable with mean 1 and S.D 3, find the probabilities that

a) $3.43 \le x \le 6.19$

b) $-1.43 \le x \le 6.19$

Sol: Given mean $\mu = 1$

Standard Deviation $\sigma = 3$

(a) If $x = 3.43$

$$Z = \frac{x-\mu}{\sigma}$$

$$= \frac{3.43-1}{3}$$

$$Z = 0.81$$

$$= Z_1$$

If $x = 6.19$

$$Z = \frac{x-\mu}{\sigma}$$

$$= \frac{6.19-1}{3}$$

$$Z = 1.73$$

$$= Z_2$$

Here Z_1 and Z_2 are of same signs then

$$P(Z_1 \le Z \le Z_2) = |A(Z_2) - A(Z_1)|$$

$$= |A(1.73) - A(0.81)|$$

$$= |0.4582 - 0.2910|$$

$$= 0.1672$$

(b) If $x = -1.43$

$$Z = \frac{x-\mu}{\sigma}$$

$$= \frac{-1.43-1}{3}$$

$$= -0.81$$

$$= Z_1$$

If $x = 6.19$

$$Z = \frac{x-\mu}{\sigma}$$

$$Z = \frac{6.19-1}{3}$$

$$Z = 1.73$$

$$= Z_2$$

Here Z_1 and Z_2 are of opposite sign then

$$P(Z_1 \le Z \le Z_2) = A(Z_2) + A(Z_1)$$

$$= A(1.73) + A(-0.81)$$

$$= A(1.73) + A(0.81)$$

$$= 0.4582 + 0.2910$$

$$= 0.7492$$

(2) If X is a normal variate with mean 30 and S.D 5

Find

a) $P(26 \le x \le 40)$

b) $P(x \ge 45)$

$\mu = 30$

$\sigma = 5$

Sol: a) If $x = 26$

$$z = \frac{x-\mu}{\sigma}$$

$$= \frac{26-30}{5}$$

$$= -0.8$$

$$= z_1$$

If $x = 40$

$$z = \frac{x-\mu}{\sigma}$$

$$= \frac{40-30}{5}$$

$$= 2$$

$$= z_2$$

z_1 and z_2 are of opposite signs then

$$P(z_1 \le z \le z_2) = A(z_2) + A(z_1)$$

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

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

$$= 0.4772 + 0.2881$$

$$= 0.7653$$

b) If $x = 45$

$$z = \frac{x-\mu}{\sigma}$$

$$= \frac{45-30}{5}$$

$$= 3$$

$$= z_1 > 0$$

$P(X \geq 45) = 0.5 - A(z_1)$

$= 0.5 - A(3)$

$= 0.5 - 0.49987$

$= 0.0013$

3) The mean and S.D of a normal variable are 8 and 4 respectively. Find

a) $P(5 \leq x \leq 10)$

b) $P(x \geq 5)$

<u>Sol</u>: a) $\mu = 8$, $\sigma = 4$

If $x = 5$

$z = \frac{x-\mu}{\sigma}$

$= \frac{5-8}{4}$

$= -0.75$

$= z_1$

If $x = 10$

$z = \frac{x-\mu}{\sigma}$

$= \frac{10-8}{4}$

$= 0.5$

$= z_2$

z_1 and z_2 are of opposite signs

$P(z_1 \leq z \leq z_2) = A(z_2) + A(z_1)$

$= A(0.5) + A(-0.75)$

$= A(0.5) + A(0.75)$

$= 0.1916 + 0.2734$

$= 0.465$

(b) $P(x \geq 5)$

If $x = 5$

$$z = \frac{x-\mu}{\sigma}$$

$$= \frac{5-8}{4}$$

$$= -0.75$$

$$= z_1 < 0$$

$P(x \geq 5)$ and $z_1 < 0$

$$P(x \geq 5) = 0.5 + A(z_1)$$

$$= 0.5 + A(-0.75)$$

$$= 0.5 + A(0.75)$$

$$= 0.5 + 0.2734$$

$$= 0.7734$$

26/10/21

(4) The mean and S.D of the marks obtained by 1000 students in an examination are respectively 34.5 and 16.5. Assuming the distribution is normal, find the approximate no. of students expected to obtain marks between 30 and 60.

Sol: Given mean $\mu = 34.5$ and S.D $\sigma = 16.5$

If $x = 30$

$$z = \frac{x-\mu}{\sigma}$$

$$= \frac{30-34.5}{16.5} = -0.27 = z_1$$

If $x = 60$

$$Z = \frac{x-\mu}{\sigma}$$

$$= \frac{60-34.5}{16.5}$$

$$= 1.54 = Z_2$$

Z_1 and Z_2 are of opposite signs

$$P(30 \leq Z \leq 60) = A(Z_2) + A(Z_1)$$

$$= A(1.54) + A(-0.27)$$

$$= 0.4382 + 0.1084$$

$$= 0.5466$$

$\therefore$ The no. of students who get marks between 30 and 60 $= 1000 \times 0.5466$

$$= 546.6$$

$$= 547$$

ESSAY

⑤ In a sample of 1000 cases the mean of a certain test is 14 and S.D is 2.5. Assuming the distribution to be normal find

a) How many students score b/w 12 and 15

b) How many score above 18

c) How many score below 18

Sol: Given $\mu = 14$ $\sigma = 2.5$

i) If $x = 12$

$$Z = \frac{x-\mu}{\sigma}$$

$$= \frac{12-14}{2.5}$$

$$= -0.8 = Z_1$$

If $x = 15$

$$Z = \frac{x-\mu}{\sigma}$$

$$= \frac{15-14}{2.5}$$

$$= 0.4$$

$$= Z_2$$

z_1 and z_2 are of opposite signs

$P(12 \le z \le 15) = A(z_2) + A(z_1)$

$= A(0.4) + A(-0.8)$

$= A(0.4) + A(0.8)$

$= 0.1554 + 0.2881$

$= 0.4435$

The no. of students who scored between 12 and 15

is $= 0.4435 \times 1000$

$= 443.5$

$= 443$

(b) $P(z > 18)$

If $x = 18$

$z = \frac{x - \mu}{\sigma}$

$= \frac{18 - 14}{2.5}$

$= 1.6$

$= z_1 > 0$

$P(z \geq 18) = 0.5 - A(z_1)$

$= 0.5 - A(1.6)$

$= 0.5 - 0.4452$

$= 0.0548$

The no. of students who scored above 18 is

0.0548×1000

54.8

③ $P(x<18)$

If $x=18$

$Z=\frac{x-\mu}{\sigma}$

$=\frac{18-14}{2.5}$

$=1.6$

$=Z_1>0$

$P(Z<18) = 0.5+A(Z_1)$

$= 0.5+A(1.6)$

$= 0.5+0.4452$

$= 0.9452$

The no. of students who scored below 18 is =

0.9452×1000

$= 945.2$

$= 945$

***

⑥ If the masses of 300 students are normally distributed with mean 68kgs and S.D 3kgs, find

a) greater than 72 kgs

b) Less than or equal to 64 kgs

c) Between 65kgs and 71 kgs

Sol: Given $\mu=68$, $\sigma=3$

ⓐ $P(Z>72)$

If $x=72$

$Z=\frac{x-\mu}{\sigma}$

$=\frac{72-68}{3}$

$=1.33 = Z_1>0$

$P(Z > 72) = 0.5 - A(z_1)$

$= 0.5 - A(1.33)$

$= 0.5 - 0.4082$

$= 0.0918$

The no. of students above 72 kgs $= 0.0918 \times 300$

$= 27.54$

$= 27$

(b) $P(Z \leq 64)$

If $x = 64$

$$z = \frac{x - \mu}{\sigma}$$

$$= \frac{64 - 68}{3}$$

$= -1.33$

$= z_1 < 0$

$P(Z \leq 64) = 0.5 - A(z_1)$

$= 0.5 - A(-1.33)$

$= 0.5 - 0.4082$

$= 0.0918$

The no. of students below 64 kgs $= 0.0918 \times 300$

$= 27.54$

$= 27$

(c) $P(65 \le z \le 71)$

If $x = 65$

$$z = \frac{x-\mu}{\sigma}$$

$$= \frac{65-68}{3}$$

$$= -1$$

$$= z_1$$

If $x = 71$

$$z = \frac{x-\mu}{\sigma}$$

$$= \frac{71-68}{3}$$

$$= 1$$

$$= z_2$$

z_1 and z_2 are of opposite signs

$$P(65 \le z \le 71) = A(z_2) + A(z_1)$$

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

$$= 0.3413 + 0.3413$$

$$= 0.6826$$

The no. of students between 65kgs and 71 kgs

$= 0.6826 \times 300$

$= 204.78$

$= 205$

*****

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

Sol: Given, 7% of items are under 35

i.e $P(X < 35) = 7\%$

$$= \frac{7}{100}$$

$$= 0.07$$

89% of items are under 63

$$P(X < 63) = 89\%$$

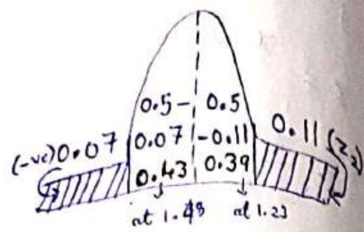
$$= \frac{89}{100} = 0.89$$

$P(x<63) = 0.89$

$P(x>63) = 1 - P(x<63)$

$= 1 - 0.89$

$P(x>63) = 0.11$



If $x = 35$

$z = \frac{x-\mu}{\sigma}$

$= \frac{35-\mu}{\sigma} = -z_1$ —①

If $x = 63$

$z = \frac{x-\mu}{\sigma}$

$= \frac{63-\mu}{\sigma}$

$\frac{63-\mu}{\sigma} = z_2$ —②

From normal distribution tables,

$z_1 = 1.48$

$z_2 = 1.23$

Substitute z_1, z_2 in ①,②

$\frac{35-\mu}{\sigma} = -1.48$ —③ $\Rightarrow 35 - \mu = -1.48\sigma$

$\frac{63-\mu}{\sigma} = 1.23$ —④

$\frac{③}{④} \Rightarrow \frac{35-\mu}{63-\mu} = \frac{-1.48}{1.23}$

$35-\mu = -1.203(63-\mu)$

$35-\mu = -75.80 + 1.20\mu$

$2.20\mu = 110.80$

$\mu = 50.36$

Mean $\mu = 50.3$

③ $\Rightarrow \dfrac{35-50.3}{\sigma} = -1.48$

$\Rightarrow +15.3 = +1.48\,\sigma$

$\Rightarrow \sigma = 10.33$

Variance $\sigma^2 = (10.33)^2$

$= 106.70$

$\therefore \mu = 50, \sigma = 10$

3) In a normal distribution, 31% of items are under 45 and 8% are over 64. Find the mean & variance of the distribution.

Sol: Given, 31% of items are under 45

$P(x<45) = 31\%$

$= \dfrac{31}{100}$

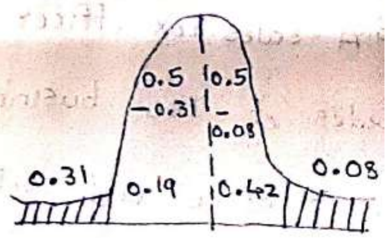
$= 0.31$

8% of items are over 64

$P(x>64) = 8\%$

$= \dfrac{8}{100}$

$= 0.08$



If $x = 45$

$z = \dfrac{x-\mu}{\sigma}$

$= \dfrac{45-\mu}{\sigma} = -z_1$ —①

If $x = 64$

$z = \dfrac{x-\mu}{\sigma}$

$= \dfrac{64-\mu}{\sigma} = z_2$ —②

From normal distribution tables

$z_1 = -0.5$

$z_2 = 1.41$

Substitute z_1, z_2 in ①,②

$$\frac{45-\mu}{\sigma} = -0.5 \Rightarrow 45-\mu = -0.5\sigma \quad —③$$

$$\frac{64-\mu}{\sigma} = 1.41 \Rightarrow 64-\mu = 1.41\sigma \quad —④$$

$$③-④ \Rightarrow 45-\cancel{\mu}-64+\cancel{\mu} = -0.5\sigma - 1.41\sigma$$

$$\cancel{-}19 = \cancel{+}1.91\sigma$$

$$\sigma = 9.9$$

$$\therefore \sigma = 10$$

$$\therefore \text{Variance } \sigma^2 = 100$$

$$③ \Rightarrow 45-\mu = -0.5(10) \Rightarrow 45-\mu = -5$$

$$\Rightarrow \therefore \mu = 50$$

H.W

⑨ A sales tax officer has reported that the average sales of 500 business he has to deal during a year is ₹ 36000 and S.D = 10000, find

a) The no. of business as the sales of which are ₹ 40000

b) The percentage of business of the sales are likely to range between ₹ 30000 and ₹ 40000

(16) 1000 students have written an examination. The mean of the test is 35 and S.D is 5, assuming the distribution to be normal, find

a) How many students marks lie between 25 & 40

b) How many students get more than 40

c) How many students get below 20

)

Sol: $\mu = 36000$ S.D $\sigma = 10000$

a) If $X = 40000$

$$Z = \frac{x - \mu}{\sigma}$$

$$= \frac{40000 - 36000}{10000}$$

$$= \frac{4000}{}$$

(b) If $x = 30000$

$z = \frac{x-\mu}{\sigma}$

$= \frac{30000-36000}{10000}$

$= \frac{-6000}{10000}$

$= -0.6$

$= z_1$

If $x = 40000$

$z = \frac{x-\mu}{\sigma}$

$= \frac{40000-36000}{10000}$

$= \frac{4000}{10000}$

$= 0.4$

$= z_2$

z_1 and z_2 are of opposite signs

$P(30000 < x < 40000) = A(z_2) + A(z_1)$

$= A(0.4) + A(-0.6)$

$= A(0.4) + A(0.6)$

$= 0.1554 + 0.2258$

$= 0.3812$

The percentage of business $= 0.3812 \times 100$

$= 38.12\%$

(10)

Sol: Given $\mu = 35$, $\sigma = 5$

(a) If $x = 25$

$z = \frac{x-\mu}{\sigma}$

$= \frac{25-35}{5}$

$= -2$

$= z_1$

If $x = 40$

$z = \frac{40-35}{5}$

$= 1$

$= z_2$

z_1 and z_2 are of opposite signs

$P(25<x<40) = A(z_2) + A(z_1)$

$= A(1) + A(-2)$

$= 0.3413 + 0.4772$

$= 0.8185$

$\therefore$ No. of students between 25 and 40 is $= 1000 \times 0.8185$

$= 818.5$

$= 819$

(b) If $x=40$

$z = \frac{x-\mu}{\sigma}$

$= \frac{40-35}{5}$

$= 1 > 0$

$P(x>40) = 0.5 - A(z_1)$

$= 0.5 - A(1)$

$= 0.5 - 0.3413$

$= 0.1587$

$\therefore$ No. of students greater than 40 $= 1000 \times 0.1587$

$= 158.7$

$= 159$

(c) If $x=20$

$z = \frac{x-\mu}{\sigma}$

$= \frac{20-35}{5}$

$= -3 < 0$

$$P(X<20) = 0.5 - A(z_1)$$

$$= 0.5 - A(-3)$$

$$= 0.5 - A(3)$$

$$= 0.5 - 0.4987$$

$$= 0.0013$$

$\therefore$ No. of students below 20 is $= 0.0013 \times 1000$

$= 1.3$

$= 1$

$\boxed{27/10/21}$

Normal Approximation to Binomial Distribution:

The normal distribution can be used to approximate the B.D. Suppose the no. of success x ranges from x_1 to x_2. Then the probability of getting x_1 to x_2 success is given by

$$\sum_{r=x_1}^{x_2} {}^nC_x\, p^x q^{n-x}$$

* If n is large the calculation of binomial Probabilities is very difficult. In such cases, the binomial curve can be replaced by normal curve and the required probability is calculated by using he following formula

For any success x, the real class interval is
$\left(x-\frac{1}{2}, x+\frac{1}{2}\right)$

Hence Z_1 corresponds to lower limit of x_1 and Z_2 corresponds upper limit of x_2 i.e

$$Z_1 = \frac{\left(x_1-\frac{1}{2}\right)-\mu}{\sigma}$$

$$Z_2 = \frac{\left(x_2+\frac{1}{2}\right)-\mu}{\sigma}$$

where μ, σ are mean, S.D of Binomial Distribution

Mean $\mu = np$

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

Eg: ① Find the probability that out of 100 patients between 84 and 95 will survive a heart operation given that the chance of survival is 0.9

Sol:- Given $p = 0.9$ $\quad n = 100$

$q = 1-0.9$
$= 0.1$

By Normal Approximation to Binomial Distribution, we have

$$Z_1 = \frac{\left(x_1-\frac{1}{2}\right)-\mu}{\sigma}$$

$$Z_2 = \frac{\left(x_2+\frac{1}{2}\right)-\mu}{\sigma}$$

Here $x_1 = 84$ and $x_2 = 95$

Mean $\mu = np$
$= 100(0.9)$
$= 90$

Variance $\sigma^2 = npq$

$= 100(0.9)(0.1)$

$= 9$

S.D $\sigma = \sqrt{9} = 3$

$$z_1 = \frac{(x_1 - \frac{1}{2}) - \mu}{\sigma} = \frac{(84 - \frac{1}{2}) - 90}{3} = \frac{-6 - \frac{1}{2}}{3}$$

$$= \frac{-13}{6}$$

$$z_1 = \cancel{-4.33}\ -2.16$$

$$z_2 = \frac{(x_2 + \frac{1}{2}) - \mu}{\sigma} = \frac{95 + \frac{1}{2} - 90}{3} = \frac{5 + \frac{1}{2}}{3}$$

$$= \frac{11}{6}$$

$$z_2 = 1.83$$

z_1 and z_2 are of opposite signs

$P(z_1 \le z \le z_2) = A(z_2) + A(z_1)$

$= A(1.83) + A(-2.16)$

$= A(1.83) + A(2.16)$

$= 0.4664 + 0.4846$

$= 0.951$

② Find the probability that a student can correctl[y] [a]nswer 25 to 30 questions in a MCQ quiz [co]nsisting of 80 questions. Assume that in each [que]stion with 4 choices, only one choice is correct.

10): Given $n = 80$

$p = \frac{1}{4}$

$q = 1 - \frac{1}{4}$

$= \frac{3}{4}$

Mean $\mu = np$

$= 80\left(\frac{1}{4}\right)$

$= 20$

$\sigma^2 = npq$

$= \cancel{80} \times \frac{1}{\cancel{4}} \times \frac{3}{\cancel{4}}$

$= 15$

$S.D = \sqrt{15} = 3.87$

$$z_1 = \frac{\left(x_1 - \frac{1}{2}\right) - \mu}{\sigma} = \frac{25 - \frac{1}{2} - 20}{3.87} = \frac{5 - \frac{1}{2}}{3.87}$$

$$= \frac{9}{(3.87)2}$$

$$= 1.16$$

Normal Distribution:

(11)

If X is a normal variable, find the area

a) to the left of $z = -1.78$

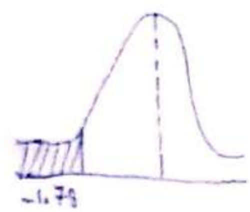
b) to the right of $z = -1.45$

c) corresponding to $-0.9 \leq z \leq 1.53$

d) to the left of $z = -2.52$ and to the right of $z = 1.83$

Sol:

a) left of $z = -1.78$



Required area $= 0.5 - A(\text{from } 0 \text{ to } -1.78)$

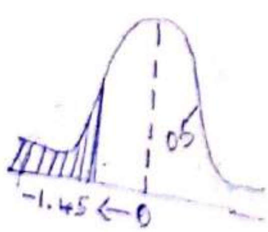
$= 0.5 - A(-1.78)$

$= 0.5 - A(1.78)$

$= 0.5 - 0.4625$

$= 0.0375$

b) to the right of $z = -1.45$



$A(0 \text{ to } -1.45)$.

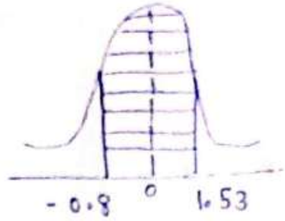
Required area $= 0.5 + A(-1.45)$

$= 0.5 + A(1.45)$

$= 0.5 + 0.4265$

$= 0.9265$

c) corresponding to $-0.8 \le z \le 1.53$



Required area =

If $x = -0.8$

$$z = \frac{x-\mu}{\sigma}$$

$$= \frac{-0.8-\mu}{\sigma} = z_1 \quad -(1)$$

If $x = 1.53$

$$z = \frac{x-\mu}{\sigma}$$

$$= \frac{1.53-\mu}{\sigma} = z_2 \quad -(2)$$

Required area = A(from 0 to -0.8) + A(from 0 to 1.53)

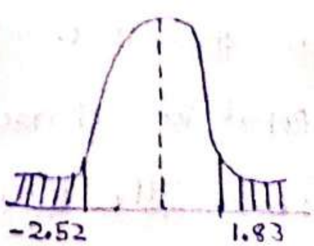
$= A(-0.8) + A(1.53)$

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

$= 0.2881 + 0.4370$

$= 0.7251$

d) left of $z = -2.52$ and right of $z = 1.83$



Required area $= 0.5 - A(0 \text{ to } -2.52) + 0.5 - A(0 \text{ to } 1.83)$

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

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

$= [0.5 - 0.4941] + [0.5 - 0.4664]$

$= 0.0395$

Sampling Distributions:

Population: Population or Universe is the aggregate or totality of statistical data forming the subject of investigation.

For example, (1) the population of the heights of Indians

(2) The population of nationalised banks in India

Definition: The no. of observations in the population is defined to be the size of the population. It may be finite or infinite. Size of the population is denoted by "N"

* Most of the times, study of entire population may not be possible to carry out and hence a part is selected from the given population.

* A portion of the population which is examined wit a view to determine the population characteristic s called a Sample. The size of the sample is noted by "n".

01/11/21

Sampling: 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 by taking into the hand from the bag and then decide to purchase it or not. Here the bag of rice is considered as population and the portion taken from the bag is considered as Sample.

Classification of Samples: Samples are classified into two ways

a) Large Sample → If the size of the sample $n > 30$, the sample is said to be large sample.

b) Small Sample → If the size of the sample $n \leq 30$, the sample is said to be small sample or simple sample.

Parameter: Any statistical measure computed from the population data is known as Parameter.

Eg: The population mean μ, Variance σ^2 are known as Parameters.

Statistic: Any Statistical measure computed from Sample data is known as Statistic.

Eg: Sample Mean $\bar{x}$, Sample Variance s^2 are known as Statistic

Formulae:

① The correction factor is

$$C.F = \frac{N-n}{N-1}$$

* There are two types of samplings

a) With replacement b) Without replacement

* The total no. of samples with replacement is given by N^n.

* The total no. of samples without replacement is given by ${}^{N}C_{n}$.

*** SAQ

Eg: ① What is the value of correction factor if $n=5$, $N=200$

Sol: Given size of the population $N=200$

Size of the sample $n=5$

$$\text{Correction Factor} = \frac{N-n}{N-1}$$

$$= \frac{200-5}{200-1}$$

$$= \frac{195}{199}$$

$$= 0.9799$$

2) A population consist of 5 numbers 2,3,6,8 and 11. Consider all possible samples of size 2 which can be drawn i) With replacement ii) Without replacement from the population. Find

i) The mean of the population

ii) The Standard Deviation of the population

iii) The mean of the sampling distribution of mean

iv) The S.D of the sampling distribution of mean

Sol: Given Size of the population $N = 5$

Size of the sample $n=2$

i) Mean of Population $\mu = \dfrac{\sum_{i=1}^{n} x_i}{n}$

$$= \frac{2+3+6+8+11}{5}$$

$$= \frac{30}{5}$$

$$\mu = 6$$

ii) Standard Deviation of Population

Variance $\sigma^2 = \dfrac{\sum(x_i - \bar{x})^2}{n}$

$$= \frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5}$$

$$= \frac{16+9+0+4+25}{5}$$

$$= \frac{54}{5}$$

$$\sigma^2 = 10.8$$

$S.D = \sqrt{10.8}$

$= 3.286$

iii) <u>With replacement:</u>

The total no. of samples with replacement is given by

$N^n = 5^2$

$= 25$ samples

The samples are

$$\left\{\begin{matrix} (2,2) & (2,3) & (2,6) & (2,8) & (2,11) \\ (3,2) & (3,3) & (3,6) & (3,8) & (3,11) \\ (6,2) & (6,3) & (6,6) & (6,8) & (6,11) \\ (8,2) & (8,3) & (8,6) & (8,8) & (8,11) \\ (11,2) & (11,3) & (11,6) & (11,8) & (11,11) \end{matrix}\right\}$$

Find the mean of each sample in the above set

$$\left\{\begin{matrix} 2 & 2.5 & 4 & 5 & 6.5 \\ 2.5 & 3 & 4.5 & 5.5 & 7 \\ 4 & 4.5 & 6 & 7 & 8.5 \\ 5 & 5.5 & 7 & 8 & 9.5 \\ 6.5 & 7 & 8.5 & 9.5 & 11 \end{matrix}\right\} \longrightarrow \text{I}$$

Mean of Sampling Distribution of Mean

$$\mu_{\bar{x}} = \frac{(2+2.5+4+5+6.5)+(2.5+3+4.5+5.5+7)+(4+4.5+6+7+8.5)+(5+5.5+7+8+9.5)+(6.5+7+8.5+9.5+11)}{25}$$

$$\mu_{\bar{x}} = \frac{150}{25} = 6$$

iv) Variance of Sampling distribution of mean:

$$\sigma_{\bar{x}}^2 = \frac{\sum (\bar{x}_i - \bar{\bar{x}})^2}{n}$$

$$= \frac{\cancel{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2}}{25}$$

$$= \cancel{16 + 9 + 0 + 4 +}$$

$$= \frac{\begin{array}{l}(2-6)^2 + (2.5-6)^2 + (4-6)^2 + (6.5-6)^2 + (5-6)^2 + (2.5-6)^2 + (3-6)^2 + (4.5-6)^2 \\ + (5.5-6)^2 + (7-6)^2 + (4-6)^2 + (4.5-6)^2 + (6-6)^2 + (7-6)^2 + (8.5-6)^2 + (5-6)^2 + \\ (5.5-6)^2 + (7-6)^2 + (8-6)^2 + (9.5-6)^2 + (6.5-6)^2 + (7-6)^2 + (8.5-6)^2 + (9.5-6)^2 \\ + (11-6)^2\end{array}}{25}$$

$$\sigma_{\bar{x}}^2 = 5.40$$

S.D $\sigma_{\bar{x}} = \sqrt{5.40}$

$= 2.32$

<u>Without Replacement:</u>

The total no. of samples without replacement is given by $^5C_2 = \frac{5\times4}{2} = 10$

The 10 Samples are

$$\left\{\begin{array}{cccc} (2,3) & (2,6) & (2,8) & (2,11) \\ & (3,6) & (3,8) & (3,11) \\ & & (6,8) & (6,11) \\ & & & (8,11) \end{array}\right\}$$

Find the mean of above sample

$$\left\{ \begin{array}{cccc} 2.5 & 4 & 5 & 6.5 \\ & 4.5 & 5.5 & 7 \\ & & 7 & 8.5 \\ & & & 9.5 \end{array} \right\}$$

$$\mu_{\bar{x}} = \frac{2.5+4+5+6.5+4.5+5.5+7+7+8.5+9.5}{10}$$

$$= \frac{60}{10}$$

$$= 6$$

Variance of Sampling Distribution of Mean

$$\sigma_{\bar{x}}^2 = \sum \frac{(x_i - \bar{x})^2}{N}$$

$$= \frac{(2.5-6)^2 + (4-6)^2 + (5-6)^2 + (6.5-6)^2 + (4.5-6)^2 + (5.5-6)^2 + (7-6)^2 + (7-6)^2 + (8.5-6)^2 + (9.5-6)^2}{10}$$

$$= \frac{40.5}{10}$$

$$\sigma_{\bar{x}}^2 = 4.05$$

$$S.D = \sqrt{4.05}$$

$$\doteq 2.012$$

③ A population consist of 5, 10, 14, 18, 13, 24 Consider all samples of size 2 which can be drawn without replacement. Find

i) The mean of the population

ii) The S.D of the population

iii) The mean of the sampling distribution of mean

iv) The S.D of the sampling distribution of mean

Sol: Given Size of the population $N=6$

Size of the sample $n=2$

i) Mean of the population $= \dfrac{\sum_{i=1}^{n} x_i}{N}$

$$= \frac{5+10+14+18+13+24}{6}$$

$$\mu = \frac{84}{6} = 14$$

ii) Variance $\sigma^2 = \dfrac{\sum(x_i-\bar{x})^2}{n}$

$$= \frac{(5-14)^2+(10-14)^2+(14-14)^2+(18-14)^2+(13-14)^2+(24-14)^2}{6}$$

$$= \frac{81+16+0+16+1+100}{6} = \frac{214}{6}$$

$$\sigma^2 = 35.6$$

$$S.D = \sqrt{35.6}$$

$$= 5.96$$

iii) The total number of samples without replacement is given by $^6C_2 = \dfrac{6\times5}{2} = 15$

$$\left\{\begin{array}{ccccc} (5,10) & (5,14) & (5,18) & (5,13) & (5,24) \\ & (10,14) & (10,18) & (10,13) & (10,24) \\ & & (14,18) & (14,13) & (14,24) \\ & & & (18,13) & (18,24) \\ & & & & (13,24) \end{array}\right\} \text{— (I)}$$

Find the mean of above sample

$$\begin{array}{ccccc} 7.5 & 9.5 & 11.5 & 9 & 14.5 \\ & 12 & 14 & 11.5 & 17 \\ & & 16 & 13.5 & 19 \\ & & & 15.5 & 21 \\ & & & & 19.5 \end{array}$$

Mean of the sampling distribution of mean

$$\mu_{\bar{x}} = \frac{7.5+9.5+11.5+9+14.5+12+14+11.5+17+16+13.5+19+15.5 + 21+18.5}{15}$$

$$= \frac{210}{15}$$

$$\mu_{\bar{x}} = 14$$

iv) Variance of Sampling distribution of mean

$$\sigma_{\bar{x}}^2 = \sum \frac{(x_i-\bar{x})^2}{N}$$

$$= \frac{\begin{array}{l}(7.5-14)^2+(9.5-14)^2+(11.5-14)^2+(9-14)^2+(14.5-14)^2+(12-14)^2+(14-14)^2+(11.5-14)^2 \\ +(17-14)^2+(16-14)^2+(13.5-14)^2+(19-14)^2+(15.5-14)^2+(21-14)^2+(18.5-14)^2\end{array}}{15}$$

$$= \frac{214}{15}$$

$$= 14.266$$

Standard Deviation $\sigma_{\bar{x}} = \sqrt{14.26} = 3.78$

④ Samples of size 2 are taken from the population

3, 6, 9, 15, 27 with replacement. Find

i) The mean of the population

ii) The S.D of the population

iii) The mean of the sampling distribution of mean

iv) The S.D of the sampling distribution of mean

Sol: Given size of population $N = 5$

Size of sample $n = 2$,

i) Mean of the population $= \dfrac{\sum_{i=1}^{n} x_i}{N}$

$$= \frac{3+6+9+15+27}{5}$$

$$= \frac{60}{5}$$

$$\mu = 12$$

ii) Standard Deviation of the population $\sigma^2 = \dfrac{\sum(x_i - \bar{x})^2}{N}$

$$= \frac{(3-12)^2 + (6-12)^2 + (9-12)^2 + (15-12)^2 + (27-12)^2}{5}$$

$$= \frac{81+36+9+9+225}{5}$$

$$= \frac{360}{5} = 72$$

$$S.D = \sqrt{72}$$

$$= 8.485$$

iii) The total number of samples with replacement is

$= 5^2 = 25$

The samples are

$$\left\{\begin{matrix} (3,3) & (3,6) & (3,9) & (3,15) & (3,27) \\ (6,3) & (6,6) & (6,9) & (6,15) & (6,27) \\ (9,3) & (9,6) & (9,9) & (9,15) & (9,27) \\ (15,3) & (15,6) & (15,9) & (15,15) & (15,27) \\ (27,3) & (27,6) & (27,9) & (27,15) & (27,27) \end{matrix}\right\} \quad (1)$$

Mean of above sample

$$\left\{\begin{matrix} 3 & 4.5 & 6 & 9 & 15 \\ 4.5 & 6 & 7.5 & 10.5 & 16.5 \\ 6 & 7.5 & 9 & 12 & 18 \\ 9 & 10.5 & 12 & 15 & 21 \\ 15 & 16.5 & 18 & 21 & 27 \end{matrix}\right\}$$

Mean of sampling distribution of mean

$$\mu_{\bar{x}} = \frac{3+4.5+6+9+15+4.5+6+7.5+10.5+16.5+6+7.5+9+12+18+9+10.5+12+15+21+15+16.5+18+21+27}{25}$$

$$= \frac{300}{25}$$

$$\mu_{\bar{x}} = 12$$

iv) Variance of the sampling distribution of mean

$$\sigma_{\bar{x}}^2 = \frac{\left[(3-12)^2+(4.5-12)^2+(6-12)^2+(9-12)^2+(15-12)^2+(4.5-12)^2+(6-12)^2+(7.5-12)^2+(10.5-12)^2+(16.5-12)^2+(6-12)^2+(7.5-12)^2+(9-12)^2+(12-12)^2+(18-12)^2+(9-12)^2+(10.5-12)^2+(12-12)^2+(15-12)^2+(21-12)^2+(15-12)^2+(16.5-12)^2+(18-12)^2+(21-12)^2+(27-12)^2\right]}{25}$$

$$\sigma_{\bar{x}}^2 = \frac{900}{25}$$

$$= 36$$

S.D $\sigma_{\bar{x}} = \sqrt{36}$

$$= 6$$

(5) If the population is 3, 6, 9, 15, 27.

i) List all possible samples of size 3 that can be taken without replacement from the given population. Find

a) The mean of the population

b) S.D of the population

c) Mean of the sampling distribution of mean

d) S.D of the sampling distribution of mean

Sol: Size of the population $N=5$

Size of the sample $n=3$

a) Mean $\mu = \frac{\sum_{i=1}^{n} x_i}{n}$

$$= \frac{3+6+9+15+27}{5}$$

$$= \frac{60}{5}$$

$$\mu = 12$$

b) S.D of the population

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

$$= \frac{(3-12)^2 + (6-12)^2 + (9-12)^2 + (15-12)^2 + (27-12)^2}{5}$$

$$= \frac{81+36+9+9+225}{5} = \frac{360}{5}$$

$$\sigma^2 = 72$$

S.D $\sigma = \sqrt{72}$

$= 8.4853$

<u>Without Replacement</u>

iii) c)
The total no. of samples without replacement is given by ${}^{N}C_{n} = {}^{5}C_{3}$

$= 10$ samples

The list of samples are

$\left\{ \begin{matrix} (3,6,9) & (3,6,15) & (3,9,15) & (3,6,27) & (3,9,27) \\ (3,15,27) & (6,9,15) & (6,9,27) & (6,15,27) & (9,15,27) \end{matrix} \right\}$

Find the mean of each sample

$\left\{ \begin{matrix} 6 & 8 & 9 & 12 & 13 \\ 15 & 10 & 14 & 16 & 17 \end{matrix} \right\}$

Mean of sampling distribution of mean

$$\mu_{\bar{x}} = \frac{6+8+9+12+13+15+10+14+16+17}{10}$$

$$= \frac{120}{10}$$

$$\mu_{\bar{x}} = 12$$

d) S.D of the sampling distribution of mean

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

$$= \frac{(6-12)^2+(8-12)^2+(9-12)^2+(12-12)^2+(13-12)^2+(15-12)^2+(10-12)^2+(14-12)^2+(16-12)^2+(17-12)^2}{10}$$

$$= \frac{36+16+9+1+9+4+4+16+25}{10} = \frac{120}{10}$$

$\sigma_{\bar{x}}^2 = 12$

$\sigma_{\bar{x}} = \sqrt{12} = 3.464$

NOTE: If $\bar{x}$ be the mean of a sample of size "n" drawn from a population with mean μ and S.D σ then

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Standard Error: It is used for calculating the difference between the expected value and the observed value.

The standard error of mean is given by $S.E = \frac{\sigma}{\sqrt{n}}$

① The variance of a population is 2. The size of the sample collected from the population is 169. What is the standard error of mean.

Sol: Given

Size of the sample $n = 169$

Variance $\sigma^2 = 2 \Rightarrow$ S.D $\sigma = \sqrt{2} = 1.414$

S.E of mean $= \frac{\sigma}{\sqrt{n}}$

$= \frac{1.414}{\sqrt{169}}$

$= \frac{1.414}{13}$

$= 0.108$

② A random sample of size 100 is taken from an infinite population having the mean $\mu = 76$ and Variance $\sigma^2 = 256$. What is the probability that $\bar{x}$ will be between 75 and 78

Sol: Given, Size of sample $n = 100$

Mean $\mu = 76$

Variance $\sigma^2 = 256 \Rightarrow$ S.D $\sigma = \sqrt{256} = 16$

If $\bar{x} = 75$ $\quad z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{75 - 76}{16/\sqrt{100}} = \dfrac{-1}{1.6} = -0.625 = z_1$

If $\bar{x} = 78$ $\quad z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{78 - 76}{16/\sqrt{100}} = \dfrac{2}{1.6} = 1.25 = z_2$

$$P(75 \le z \le 78) = A(z_2) + A(z_1)$$
$$= A(1.25) + A(-0.625)$$
$$= A(1.25) + A(0.625)$$
$$= 0.3944 + 0.2324$$
$$= 0.6268$$

③ A random sample of size 64 is taken from a normal Population with $\mu = 51.4$ & $\sigma = 6.8$. What is the probability that the mean of the sample will be

a) exceeds 52.9

b) falls between 50.5 & 52.3

c) less than 50.6

Sol: Given, Size of the sample $n = 64$

Mean $\mu = 51.4$

S.D $\sigma = 6.8$

i) $P(z > 52.9)$

If $x = 52.9$

$$Z = \frac{x - \mu}{\sigma/\sqrt{n}} = \frac{52.9 - 51.4}{6.8/\sqrt{64}} = \frac{1.5}{6.8/8} = \frac{1.5}{0.85} = 1.76 = z_1 > 0$$

$$P(z > 52.9) = 0.5 - A(1.76)$$
$$= 0.5 - 0.4608$$
$$= 0.0392$$

ii) $P(50.5 < x < 52.3)$

If $x = 50.5$

$$Z = \frac{x - \mu}{\sigma/\sqrt{n}}$$
$$= \frac{50.5 - 51.4}{6.8/\sqrt{64}}$$
$$= \frac{-0.9}{0.85}$$
$$= -1.05 = z_1$$

If $x = 52.3$

$$Z = \frac{x - \mu}{\sigma/\sqrt{n}}$$
$$= \frac{52.3 - 51.4}{6.8/\sqrt{64}}$$
$$= \frac{0.9}{0.85}$$
$$= 1.05$$
$$= z_2$$

z_1 and z_2 are of opposite sign

$$P(z_1 < z < z_2) = P(50.5 < z < 52.3) = A(z_2) + A(z_1)$$
$$= A(1.05) + A(-1.05)$$
$$= A(1.05) + A(1.05)$$
$$= 2(0.3531)$$
$$= 0.7062$$

iii) $P(z<50.6)$

If $x = 50.6$

$$z = \frac{x-\mu}{\sigma/\sqrt{n}}$$

$$= \frac{50.6 - 51.4}{6.8/\sqrt{64}}$$

$$= \frac{-0.8}{0.85}$$

$$= -0.94 = z_1 < 0$$

$$P(z<50.6) = 0.5 - A(z_1)$$
$$= 0.5 - A(-0.94)$$
$$= 0.5 - A(0.94)$$
$$= 0.5 - 0.3264$$
$$= 0.1736$$

Chebyshev's Inequality: Chebyshev's inequality helps to derive bounds on probabilities when only the mean or both the mean and variance of the probability distribution are given.

Chebyshev's Theorem: If a probability distribution has mean μ and S.D σ, the probability of getting a value which deviates μ by atleast K_σ is atmost $\frac{1}{K^2}$ i.e

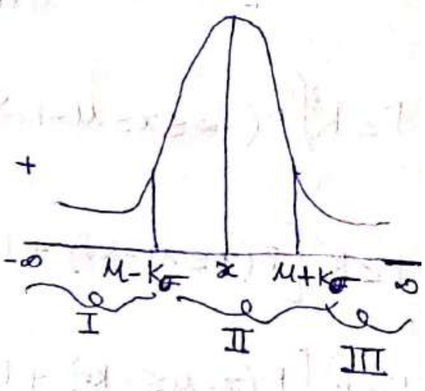
$P[|x-\mu| \geq K\sigma] \leq \frac{1}{K^2}$ is the probability assosciated ith set of outcomes for which the value of a ndom Variable having the given probability ribution is such that $|x-\mu| \geq K\sigma$

Proof: Given Mean $E(x) = \mu$ & Variance $V(x) = \sigma^2$

By definition of variance, we have

$$\sigma^2 = V(x) = E[x - E(x)]^2$$
$$= E(x-\mu)^2$$
$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 . f(x)\, dx$$

Now dividing the interval into 3 parts



From the above, we have

$$\sigma^2 = \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)dx$$

Neglecting the second integral, since second integral is always a positive value.

$$\sigma^2 \geq \underbrace{\int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\, dx}_{\text{I}} + \underbrace{\int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)dx}_{\text{II}} \quad \text{--- (1)}$$

From First Integral

$x \leq \mu - k\sigma$

$x - \mu \leq -k\sigma$

$-(x-\mu) \geq k\sigma$

$(x-\mu)^2 \geq k^2\sigma^2$ --- (2)

From Second Integral

$x \geq \mu + k\sigma$

$x - \mu \geq k\sigma$

$(x-\mu)^2 \geq k^2\sigma^2$ --- (3)

Substitute (2) & (3) in (1)

⑨ If the probability of a defective bolt is 0.2. Find

a) Mean b) Standard Deviation for the distribution of bolts in a total of 400

Sol: Given $n = 400$

$$P = \frac{2}{10} \qquad q = 1 - \frac{1}{5}$$

$$= \frac{1}{5} \qquad = \frac{4}{5}$$

Mean $\mu = np$

$$= 400\left(\frac{1}{5}\right)$$

$$= 80$$

Variance $\sigma^2 = npq$

$$= \cancel{400}^{\cancel{80}^{16}} \times \frac{1}{\cancel{5}_1} \times \frac{4}{\cancel{5}_1}$$

$$= 64$$

$$S.D = \sqrt{\text{Variance}}$$

$$= \sqrt{64}$$

$$= 8$$

⑩ The mean and variance of a binomial distribution are 4 and $\frac{4}{3}$ respectively. Find $P(x \geq 1)$

Sol: Given $\mu = np = 4$ — ①

$$\sigma^2 = npq = \frac{4}{3} \text{ — ②}$$

$$\frac{①}{②} \Rightarrow \frac{\cancel{np}}{\cancel{np}q} = \frac{\cancel{4}}{\cancel{4}} \times 3 \quad \Rightarrow \frac{1}{q} = 3$$

$\Rightarrow q = \frac{1}{3}$

$P = 1 - q$

$= 1 - \frac{1}{3}$

$= \frac{2}{3}$

$① \Rightarrow np = 4$

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

$n = 6$

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

$= 1 - {}^6C_0 \left(\frac{2}{3}\right)^0 \left(\frac{1}{3}\right)^{6-0}$

$= 1 - \frac{1}{3^6}$

$= 0.998$

(11) Determine the Binomial Distribution for which mean is 4 and Variance is 3

Sol: $\mu = np = 4$ —①

$\sigma^2 = npq = 3$ —②

$\frac{②}{①} \Rightarrow \frac{npq}{np} = \frac{3}{4} \Rightarrow q = \frac{3}{4}$

$P = 1 - q = 1 - \frac{3}{4} = \frac{1}{4}$

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

***

(12) 6 dice are thrown, 729 times. How many times do you expect atleast 3 dice to show 5 or 6

Sol: Given $n = 6$

P = Probability of getting 5 or 6

$= \frac{2}{6}$

$p = \frac{1}{3}$ $q = 1 - \frac{1}{3} = \frac{2}{3}$

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

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

$$= 1 - \left[{}^6C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^{6-0} + {}^6C_1 \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^{6-1} + {}^6C_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^{6-2} \right]$$

$$= 1 - \left[\left(\frac{2}{3}\right)^6 + \frac{6 \times 2^5}{3^6} + \frac{15 \times 2^4}{3^6} \right]$$

$$= 1 - \frac{1}{3^6} [64 + 192 + 240]$$

$$= 1 - \frac{496}{3^6}$$

$$= 0.3196$$

$$= \frac{233}{729}$$

$\therefore$ The expected no. of such cases in 729 times is $729 \times \frac{233}{729} = 233$

(13) Two dice are thrown 120 times. Find the average no. of times in which the number on the first die exceeds the number on the second die.

Sol: Given 2 dice are thrown i.e $n = 2$

Total no. of cases $= 6^2 = 36$

P = Probability of getting a number which ~~exists~~ exceeds second die

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

$= \frac{\cancel{15}\,5}{\cancel{36}\,12}$

$P = \frac{5}{12}$ $\qquad q = 1 - \frac{5}{12} = \frac{7}{12}$

Mean $\mu = np$ $\qquad n = 2$

$\boxed{\begin{aligned} &= \cancel{120}^{10} \times \frac{5}{\cancel{12}_1} \\ &= 50 \end{aligned}}$ $\qquad = 2 \times \frac{5}{12}$

$\qquad = \frac{5}{6}$

$= \frac{5}{\cancel{6}_1} \times \cancel{120}^{20}$

$= 100$

(14) Out of 800 families with 5 children each how many do 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 and girls

Sol: Given $n=5$

p = Prob. of boy $=\frac{1}{2}$

q = Prob. of girl $=\frac{1}{2}$

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

$= \frac{5\times\cancel{4}^{2}}{\cancel{2}\times 1}\times\frac{1}{2^{\cancel{5}4}}$

$= \frac{5}{16}$ → Only for one family

Hence the Probability of no. of families having 3 boys is

For 800 families, $\frac{5}{\cancel{16}_1}\times\cancel{800}^{50}$

$=250$

b) $P(x=5) = {}^5C_5\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^{5-5}$

$= 1\times\frac{1}{2^5}$

$= \frac{1}{32}$

For 800 families $= \frac{1}{32}\times 800$

$= 25$

c) $P(x=2)+P(x=3) = {}^5C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{5-2} + {}^5C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{5-3}$

$= 2\times\frac{5\times 4}{2\times 1}\left(\frac{1}{2^5}\right)$

$= \frac{\cancel{20}^{10}}{\cancel{32}_{16}}$

$= \frac{10}{16}$

For 800 families, $\frac{10}{\cancel{16}_1} \times \cancel{800}^{50}$

$= 500$

d) $P(x \geq 1) = 1 - P(x=0)$

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

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

$= \frac{31}{32}$

For 800 families, $\frac{31}{32} \times 800$

$= 775$

***

(15) Fit a Binomial Distribution to the following frequency distribution

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

Sol: Given $n = 6$

$N =$ Total frequency

$= \sum f_i$

$= 13 + 25 + 52 + 58 + 32 + 16 + 4$

$= 200$

Mean $= \frac{\sum f_i x_i}{\sum f_i} = \frac{0(13) + 1(25) + 2(52) + 3(58) + 4(32) + 5(16) + 6(4)}{200}$

$= \frac{535}{200}$

$= 2.675$

$np = 2.675 \Rightarrow P = \frac{2.675}{6}$

$\Rightarrow P = 0.445$

$q = 1 - P$

$= 1 - 0.445$

$q = 0.555$

x	f	$P(x) = {}^nC_x q^{n-x} p^x$	Expected Frequency $E = N \times P(x)$	Rounding off
0	18	$P(0) = {}^6C_0 (0.555)^6 (0.445)^0$ $= 0.029$	$= 200 \times P(0)$ $= 200 \times (0.029)$ $= 5.84$	6
1	25	$P(1) = {}^6C_1 (0.555)^5 (0.445)^1$ $= 0.14$	$= 200 \times P(1)$ $= 28.11$	28
2	52	$P(2) = {}^6C_2 (0.555)^4 (0.445)^2$ $= 0.28$	$= 200 \times P(2)$ $= 56.36$	56
3	58	$P(3) = {}^6C_3 (0.555)^3 (0.445)^3$ $= 0.30$	$= 200 \times P(3)$ $= 60.25$	60
4	32	$P(4) = {}^6C_4 (0.555)^2 (0.445)^4$ $= 0.18$	$= 200 \times P(4)$ $= 36.23$	36
5	16	$P(5) = {}^6C_5 (0.555)^1 (0.445)^5$ $= 0.058$	$= 200 \times P(5)$ $= 11.62$	12
6	4	$P(6) = {}^6C_6 (0.555)^0 (0.445)^6$ $= 7.765\ldots \times 10^{-3}$	$= 200 \times P(6)$ $= 1.55$	2

Expected Probability (Frequency) $= 6+28+56+60+36+12+2$

$= 200$

$=$ Given Probability (Frequency)

(or)

Actual Probability (Frequency)

(16) 4 coins are tossed 160 times. The no. of times head occurs is given below

x	0	1	2	3	4
No. of times	8	34	69	43	6

Fit a Binomial Distribution to this data on the hypothesis that the coins are unbiased

Sol: Given $n=4$

$\therefore$ Given coins are unbiased $\Rightarrow p=\frac{1}{2}$ $q=\frac{1}{2}$

$N = \sum f_i$

$= 8+34+69+43+6$

$= 160$

x	f	$P(x) = {}^nC_x\, q^{n-x}\, p^x$	Expected Frequency $E = N \times P(x)$	Rounding Off
0	8	$P(0) = {}^4C_0 \left(\frac{1}{2}\right)^{4-0} \left(\frac{1}{2}\right)^0 = 0.0625$	$E = 160 \times P(0) = 160 \times 0.0625 = 10$	10
1	34	$P(1) = {}^4C_1 \left(\frac{1}{2}\right)^{4-1} \left(\frac{1}{2}\right)^1 = 0.25$	$= 160 \times P(1) = 160 \times 0.25 = 40$	40
2	69	$P(2) = {}^4C_2 \left(\frac{1}{2}\right)^{4-2} \left(\frac{1}{2}\right)^2 = 0.375$	$= 160 \times P(2) = 160 \times 0.375 = 60$	60
3	43	$P(3) = {}^4C_3 \left(\frac{1}{2}\right)^{4-3} \left(\frac{1}{2}\right)^3 = 0.25$	$= 160 \times P(3) = 160 \times 0.25 = 40$	40
4	6	$P(4) = {}^4C_4 \left(\frac{1}{2}\right)^{4-4} \left(\frac{1}{2}\right)^4 = 0.0625$	$= 160 \times P(4) = 160 \times 0.0625 = 10$	10

Expected Frequency $= 10 + 40 + 60 + 40 + 10$

$= 160$

$=$ Actual Frequency

20/10/21

Poisson Distribution: Poisson Distribution is a distinct distribution. This distribution is applied when the events whose probability of occurence is very small but no. of trials are very large.

n is very large
P is very small

Definition: A random variable X is said to follow Poisson distribution if it assumes only non-negative values and its probability density function is given by

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

Eg: 1) The no. of telephone calls per min at a switch board

2) The no. of cars passing a certain point in one minute

3) The no. of printing mistakes per page in a large text book.

Conditions:

1) The no. of occurences should be a discrete variable.

2) The no. of trials 'n' should be large.

3) The no. of Printing mistakes per page in a large textbook.

3) The probability of success 'p' is very small

$np = \lambda$ is finite

Mean of Poisson Distribution: The mean of Poisson distribution is λ i.e $\lambda = np$

Proof → Let X is a discrete random variable

By definition of poisson distribution, we have

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

By definition of mean$(\mu) = \sum_{x=0}^{n} x.P(x)$

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

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

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

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

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

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

$$\mu = \lambda$$

Variance of Poisson Distribution: The variance of Poisson distribution is λ

Proof → By definition of poisson distribution, we have

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

By definition of variance

$$Var(x) = \sum_{x=0}^{n} x^2 P(x) - \mu^2$$

$$= \sum_{x=0}^{n} x^2 \frac{e^{-\lambda}\lambda^x}{x!} - \lambda^2$$

$$= \sum [x(x-1)+x] \frac{e^{-\lambda}\lambda^x}{x!} - \lambda^2$$

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

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

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

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

$$= \lambda^2 e^{-\lambda}\left[\sum_{x=2}^{n} \frac{\lambda^{x-2}}{(x-2)!}\right] + \lambda - \lambda^2$$

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

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

$= \bar{x}^2 + \lambda - \bar{x}^2$

$var(x) = \lambda$

① If the probability that a person suffers a bad reaction from a certain injection is 0.001, determine the probability that out of 2000 members

a) Exactly 3

b) More than 2 persons

c) None

d) More than 1 person suffers a bad reaction

Sol: Given $p = 0.001$ $n = 2000$

By definition of poisson distribution,

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

We have $\lambda = np$

$= 2000(0.001)$

$= 2$

a) $P(x=3) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$

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

$= 0.1804$

b) $P(x>2) = 1 - P(x\leq 2)$

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

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

$= 1 - [0.135 + 0.270 + 0.270]$

$= 0.325$

c) $P(x=0) = \frac{e^{-2}2^0}{0!}$

$= 0.135$

d) $P(x>1) = 1 - P(x\leq 1)$

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

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

$= 1 - [0.135 + 0.270]$

$= 0.595$

② A manufacturer knows that condensers he makes contain an average 1% defectives. He packs them in boxes of 100. What is the probability that a box picked at random will contain 3 or more defective condensers.

Sol: Given $n = 100$ $\quad P = 1\%$

$= \frac{1}{100} = 0.01$

$\lambda = np$

$= 100(0.01)$

$= 1$

$P(x \geq 3) = 1 - P(x < 3)$

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

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

$= 1 - [0.367 + 0.367 + 0.183]$

$= 0.083$

③ A hospital switch board receives an average of 4 emergency calls in 10 min, what is the probability that

a) There are atmost 2 emergency calls

b) There are exactly 3 emergency calls

Sol: $\lambda = 4$

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

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

$= 0.0183 + 0.073 + 0.146$

$= 0.237$

b) $P(x=3) = \dfrac{e^{-4}4^3}{3!}$

$= 0.195$

④ Average no. of accidents on any day on a natural highway is 1.8. Determine the probability that the no. of accidents are a) atleast one

b) atmost one

Sol: $\lambda = 1.8$

a) $P(x \geq 1) = 1 - P(x=0)$

$= 1 - \left[\dfrac{e^{-1.8}(1.8)^0}{0!}\right]$

$= 1 - 0.165$

$= 0.835$

b) $P(x \leq 1) = P(x=0) + P(x=1)$

$= 0.165 + \dfrac{e^{-1.8}(1.8)^1}{1!}$

$= 0.165 + 0.297$

$= 0.462$

5) If the variance of a poisson distribution is 3. Find the probability that a) $x=0$

b) $0 < x \leq 3$

c) $1 \leq x \leq 4$

Sol: $\lambda = 3$

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

$= 0.049$

b) $P(0 < x \leq 3) = P(x=1) + P(x=2) + P(x=3)$

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

$= 0.149 + 0.224 + 0.224$

$= 0.597$

c) $P(1 \leq x \leq 4) = P(x=1) + P(x=2) + P(x=3) + P(x=4)$

$= 0.149 + 0.224 + 0.224 + 0.168$

$= 0.765$

23/10/21

6) 2% of items of a factory are defective. The items are packed in boxes. What is the probability that there will be

a) 2 defective items

b) Atleast 3 defective items in a box of 100 items

Sol: $n = 100$ $\quad p = 2\%$

$= \frac{2}{100}$

$\lambda = np$

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

$\lambda = 2$

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

$= 2e^{-2}$

$= 0.2706$

b) $P(x \geq 3) = 1 - P(x<3)$

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

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

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

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

$= 0.3233$

NOTE: The recurrence formula for Poisson Distribution is

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

7) If X is a Poisson variable such that $P(x=0) = P(x=1)$, find $P(x=0)$ and using recurrence formula find the probabilities at $x = 1, 2, 3, 4$ & 5

Sol: Given $P(x=0)=P(x=1)$

$$\frac{e^{-\lambda}\lambda^0}{0!}=\frac{e^{-\lambda}\lambda^1}{1!}$$

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

$$\Rightarrow \lambda=1$$

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

$$=1.e^{-1}$$

$$=0.3678$$

$\left[\because \text{By recurrence } P(x+1)=\frac{\lambda}{x+1}P(x)\right]$

$$P(x=1)=P(x+1) \quad x=0$$

$$=\frac{1}{0+1}\cdot 0.3678$$

$$=0.3678$$

$$P(x=2)=\frac{1}{2}(0.3678) \qquad P(x=2)=\frac{1}{2}P(x=1)$$

$$=0.1839$$

$$P(x=3)=\frac{1}{3}(0.1839) \qquad P(x=3)=\frac{1}{3}P(x=2)$$

$$=0.0613$$

$$P(x=4)=\frac{1}{4}(0.0613) \qquad P(x=4)=\frac{1}{4}P(x=3)$$

$$=0.015325$$

$$P(x=5)=\frac{1}{5}(0.015325) \qquad P(x=5)=\frac{1}{5}P(x=4)$$

$$=0.003065 \quad =0.0030$$

⑧ Using recurrence formula, find the probabilities when $x = 0, 1, 2, 3, 4, 5$ if the mean of P.D is 3

Sol: $\mu = \lambda = 3$

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

$$= 0.0497$$

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

$$= 3(0.0497)$$

$$= 0.1491$$

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

$$= \frac{3}{2}(0.1491)$$

$$= 0.2236$$

$$P(x=3) = \frac{3}{3}(0.2236)$$

$$= 0.2236$$

$$P(x=4) = \frac{3}{4} P(x=3)$$

$$= \frac{3}{4}(0.2236)$$

$$= 0.1677$$

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

$$= \frac{3}{5}(0.1677)$$

$$= 0.100$$

(9) If a P.D is such that $P(x=1)\cdot\frac{3}{2} = P(x=3)$

Find a) $P(x \geq 1)$

b) $P(x \leq 3)$

c) $P(2 \leq x \leq 5)$

<u>Sol:</u> $\frac{3}{2}P(x=1) = P(x=3)$

$$\frac{3}{2}\left[\frac{e^{-\lambda}\lambda^{1}}{1!}\right] = \frac{e^{-\lambda}\lambda^{\cancel{3}\,2}}{3!}$$

$$\Rightarrow \lambda^2 = \frac{3}{\cancel{2}_1}\times\cancel{6}^{3}$$

$$\Rightarrow \lambda^2 = 9 \Rightarrow \lambda = 3$$

(a) $P(x \geq 1) = 1 - P(x=0)$

$$= 1 - \left[\frac{e^{-3}3^{0}}{0!}\right]$$

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

$$= 0.9502$$

(b) $P(x \leq 3) = P(x=0) + P(x=1) + P(x=2) + P(x=3)$

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

$$= e^{-3} + 3e^{-3} + \frac{9}{2}(e^{-3}) + \frac{9}{2}(e^{-3})$$

$$= 4e^{-3} + 9e^{-3}$$

$$= 13e^{-3}$$

(c) $P(2 \le x \le 5) = P(x=2) + P(x=3) + P(x=4) + P(x=5)$

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

$$= \frac{9}{2}e^{-3} + \frac{9}{2}e^{-3} + \frac{27}{8}e^{-3} + \frac{81}{40}e^{-3}$$

$$= 9e^{-3} + \frac{216}{40}e^{-3}$$

$$= \frac{72}{5}e^{-3}$$

$$= 0.7169$$

(10) Fit a P.D for the following data and calculate the expected frequency

x	0	1	2	3	4
$f(x)$	109	65	22	3	1

sol: $N = \Sigma f_i$ = Total frequency

$= 109 + 65 + 22 + 3 + 1$

$= 200$

Mean $\mu = \frac{\Sigma f_i x_i}{\Sigma f_i}$

$$= \frac{0 + 65 + 44 + 9 + 4}{200}$$

$= \frac{122}{200}$

$= 0.61$

$\lambda = 0.61$

x	$f(x)$	$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$	Expected Frequency $= N \times P(x)$	Rounding off
0	109	$P(x) = \frac{e^{-0.61}\lambda^0}{0!}$ $= 0.5433$	$= 200 \times 0.5433$ $= 108.6$	109
1	65	$P(x) = \frac{e^{-0.61}(0.61)}{1!}$ $= 0.3314$	$= 200 \times 0.3314$ $= 66.2$	66
2	22	$P(x) = \frac{e^{-0.61}(0.61)^2}{2!}$ $= 0.1010$	$= 200 \times 0.1010$ $= 20.2$	20
3	3	$P(x) = \frac{e^{-0.61}(0.61)^3}{3!}$ $= 0.0205$	$= 200 \times 0.0205$ $= 4.11$	4
4	1	$P(x) = \frac{e^{-0.61}(0.61)^4}{4!}$ $= 3.134 \times 10^{-3}$	$= 200 \times 3.134 \times 10^{-3}$ $= 0.62$	1

(11) The distribution of typing mistakes commited by a typist is given below.

Assuming the distribution to be Poisson, find the expected frequencies.

x	0	1	2	3	4	5
$f(x)$	42	33	14	6	4	1

Sol: $\sum f_i = 42 + 33 + 14 + 6 + 4 + 1$

$= 100$

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i} = \frac{0 + 33 + 28 + 18 + 16 + 5}{100}$$

$= 1$

x	$f(x)$	$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$	Expected Frequency $= 100 \times P(x)$	Rounding off
0	42	$= \frac{e^{-1} 1^0}{0!}$ $= e^{-1} = 0.3678$	$= 100 \times 0.3678$ $= 36.78$	37
1	33	$= \frac{e^{-1} 1^1}{1!}$ $= 0.3678$	$= 100 \times 0.3678$ $= 36.78$	37
2	14	$= \frac{e^{-2}(1)^2}{2!}$ $= \cancel{0.0676}$ 0.1839	$= 100 \times \cancel{0.0676}$ 0.1839 $= \cancel{6.76}$ 18.39	18
3	6	$= \frac{e^{-1}(1)^3}{3!}$ $= 0.0061$	$= 100 \times 0.0061$ $=$ 6.131	6
4	4	$= \frac{e^{-1}(1)^4}{4!}$ $= 0.0153$	$= 100 \times 0.0153$ $= 1.53$	2
5	1	$= \frac{e^{-1}(1)^5}{5!}$ $= 3.06\ldots \times 10^{-3}$	$= 100 \times 3.06\ldots \times 10^{-3}$ $= 0.30$	0

25/10/21

Normal Distribution: The normal distribution is a continuous probability distribution. A continuous distribution is a distribution in which the Variate can take all values within a given range.

Eg: The height of a person, the speed of the vehicle etc

Definition: A random Variable X is said to have a normal distribution if its density function is given by

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \begin{array}{c} -\infty < x < \infty \\ -\infty < \mu < \infty \\ \sigma > 0 \end{array}$$

Standard Normal Distribution: The normal distribution with mean $\mu = 0$ and standard Deviation $\sigma = 1$ is known as Standard Normal Distribution.

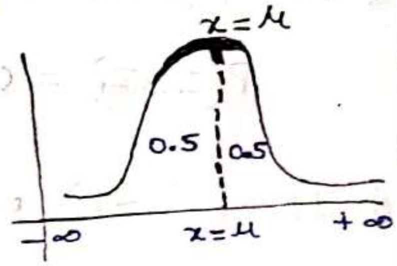
NOTE: The random Variable that follows normal distribution is denoted by Z. If a variable X follows normal distribution with mean μ and S.D σ, then the variable Z is defined as

$$Z = \frac{x - \mu}{\sigma}$$

Characteristics of Normal Distribution:

1) The graph of Normal Distribution $y=f(x)$ in the xy plane is known as Normal Curve.

2) The curve is a bell shape curve and symmetrical with respect to mean i.e about the line $x=\mu$ and the two tails on the right and left side of the mean extends to infinity



3) Area under the normal curve represents the total population.

4) Mean, Median and Mode of the distribution coincide at $x=\mu$ as the distribution is symmetrical. So the normal curve is unimodel.

5) Linear combination of independent normal variables is also a normal variable.

Formulae:

① If both z_1 & z_2 are positive (or) negative then

$$P(z_1 \leq z \leq z_2) = |A(z_2) - A(z_1)|$$ where A is area

② If both Z_1 and Z_2 are of opposite sign then

$$P(Z_1 \leq Z \leq Z_2) = A(Z_2) + A(Z_1)$$

③ If $P(Z \geq Z_1)$ or $P(Z > Z_1)$ and $Z_1 > 0$ then

$$P(Z \geq Z_1) = 0.5 - A(Z_1)$$

④ If $P(Z \geq Z_1)$ or $P(Z > Z_1)$ and $Z_1 < 0$ then

$$P(Z \geq Z_1) = 0.5 + A(Z_1)$$

⑤ If $P(Z \leq Z_1)$ or $P(Z < Z_1)$ and $Z_1 > 0$ then

$$P(Z \leq Z_1) = 0.5 + A(Z_1)$$

⑥ If $P(Z \leq Z_1)$ or $P(Z < Z_1)$ and $Z_1 < 0$ then

$$P(Z \leq Z_1) = 0.5 - A(Z_1)$$

⑦ $A(-Z_1) = A(Z_1)$ [∵ The curve is symmetric]

① For a normally distributed variable with mean 1 and S.D 3, find the probabilities that

a) $3.43 \leq x \leq 6.19$

b) $-1.43 \leq x \leq 6.19$

Sol: Given mean $\mu = 1$

Standard Deviation $\sigma = 3$

(a) If $x = 3.43$

$$Z = \frac{x-\mu}{\sigma}$$

$$= \frac{3.43-1}{3}$$

$$Z = 0.81$$

$$= Z_1$$

If $x = 6.19$

$$Z = \frac{x-\mu}{\sigma}$$

$$= \frac{6.19-1}{3}$$

$$Z = 1.73$$

$$= Z_2$$

Here Z_1 and Z_2 are of same signs then

$$P(Z_1 \le Z \le Z_2) = |A(Z_2) - A(Z_1)|$$

$$= |A(1.73) - A(0.81)|$$

$$= |0.4582 - 0.2910|$$

$$= 0.1672$$

(b) If $x = -1.43$

$$Z = \frac{x-\mu}{\sigma}$$

$$= \frac{-1.43-1}{3}$$

$$= -0.81$$

$$= Z_1$$

If $x = 6.19$

$$Z = \frac{x-\mu}{\sigma}$$

$$Z = \frac{6.19-1}{3}$$

$$Z = 1.73$$

$$= Z_2$$

Here Z_1 and Z_2 are of opposite sign then

$$P(Z_1 \le Z \le Z_2) = A(Z_2) + A(Z_1)$$

$$= A(1.73) + A(-0.81)$$

$$= A(1.73) + A(0.81)$$

$$= 0.4582 + 0.2910$$

$$= 0.7492$$

② If X is a normal variate with mean 30 and S.D 5

Find

a) $P(26 \leq x \leq 40)$

b) $P(x \geq 45)$

$\mu = 30$

$\sigma = 5$

Sol: a) If $x = 26$

$$z = \frac{x-\mu}{\sigma}$$

$$= \frac{26-30}{5}$$

$$= -0.8$$

$$= z_1$$

If $x = 40$

$$z = \frac{x-\mu}{\sigma}$$

$$= \frac{40-30}{5}$$

$$= 2$$

$$= z_2$$

z_1 and z_2 are of opposite signs then

$$P(z_1 \leq z \leq z_2) = A(z_2) + A(z_1)$$

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

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

$$= 0.4772 + 0.2881$$

$$= 0.7653$$

b) If $x = 45$

$$z = \frac{x-\mu}{\sigma}$$

$$= \frac{45-30}{5}$$

$$= 3$$

$$= z_1 > 0$$

$P(X \geq 45) = 0.5 - A(z_1)$

$= 0.5 - A(3)$

$= 0.5 - 0.49987$

$= 0.0013$

③ The mean and S.D of a normal variable are 8 and 4 respectively. Find

a) $P(5 \leq x \leq 10)$

b) $P(x \geq 5)$

Sol: a) $\mu = 8$, $\sigma = 4$

If $Z = 5$

$Z = \frac{x-\mu}{\sigma}$

$= \frac{5-8}{4}$

$= -0.75$

$= z_1$

If $x = 10$

$Z = \frac{x-\mu}{\sigma}$

$= \frac{10-8}{4}$

$= 0.5$

$= z_2$

z_1 and z_2 are of opposite signs

$P(z_1 \leq z \leq z_2) = A(z_2) + A(z_1)$

$= A(0.5) + A(-0.75)$

$= A(0.5) + A(0.75)$

$= 0.1916 + 0.2734$

$= 0.465$

(b) $P(X \geq 5)$

If $x = 5$

$$z = \frac{x-\mu}{\sigma}$$

$$= \frac{5-8}{4}$$

$$= -0.75$$

$$= z_1 < 0$$

$P(x \geq 5)$ and $z_1 < 0$

$$P(x \geq 5) = 0.5 + A(z_1)$$

$$= 0.5 + A(-0.75)$$

$$= 0.5 + A(0.75)$$

$$= 0.5 + 0.2734$$

$$= 0.7734$$

26/10/21

(4) The mean and S.D of the marks obtained by 1000 students in an examination are respectively 34.5 and 16.5. Assuming the distribution is normal, find the approximate no. of students expected to obtain marks between 30 and 60.

Sol: Given mean $\mu = 34.5$ and S.D $\sigma = 16.5$

If $x = 30$

$$Z = \frac{x-\mu}{\sigma}$$

$$= \frac{30-34.5}{16.5} = -0.27 = z_1$$

If $x = 60$

$Z = \frac{x - \mu}{\sigma}$

$= \frac{60 - 34.5}{16.5}$

$= 1.54 = Z_2$

Z_1 and Z_2 are of opposite signs

$P(30 \leq Z \leq 60) = A(Z_2) + A(Z_1)$

$= A(1.54) + A(-0.27)$

$= 0.4382 + 0.1084$

$= 0.5466$

∴ The no.of students who get marks between 30 and 60 $= 1000 \times 0.5466$

$= 546.6$

$= 547$

ESSAY

(5) In a sample of 1000 cases the mean of a certain test is 14 and S.D is 2.5. Assuming the distribution to be normal find

a) How many students score b/w 12 and 15

b) How many score above 18

c) How many score below 18

Sol: Given $\mu = 14$ $\sigma = 2.5$

a) If $x = 12$

$Z = \frac{x - \mu}{\sigma}$

$= \frac{12 - 14}{2.5}$

$= -0.8 = Z_1$

If $x = 15$

$Z = \frac{x - \mu}{\sigma}$

$= \frac{15 - 14}{2.5}$

$= 0.4$

$= Z_2$

z_1 and z_2 are of opposite signs

$$P(12 \le Z \le 15) = A(z_2) + A(z_1)$$

$$= A(0.4) + A(-0.8)$$

$$= A(0.4) + A(0.8)$$

$$= 0.1554 + 0.2881$$

$$= 0.4435$$

The no. of students who scored between 12 and 15

is $= 0.4435 \times 1000$

$= 443.5$

$= 443$

(b) $P(x > 18)$

If $x = 18$

$$z = \frac{x - \mu}{\sigma}$$

$$= \frac{18 - 14}{2.5}$$

$$= 1.6$$

$$= z_1 > 0$$

$$P(z \ge 18) = 0.5 - A(z_1)$$

$$= 0.5 - A(1.6)$$

$$= 0.5 - 0.4452$$

$$= 0.0548$$

The no. of students who scored above 18 is

$= 0.0548 \times 1000$

$= 54.8$

$= 55$

ⓒ $P(x<18)$

If $x=18$

$$z=\frac{x-\mu}{\sigma}$$

$$=\frac{18-14}{2.5}$$

$=1.6$

$=z_1>0$

$P(z<18)=0.5+A(z_1)$

$=0.5+A(1.6)$

$=0.5+0.4452$

$=0.9452$

The no. of students who scored below 18 is =

0.9452×1000

$=945.2$

$=945$

⑥ If the masses of 300 students are normally distributed with mean 68kgs and S.D 3kgs, find

a) greater than 72 kgs

b) Less than or equal to 64 kgs

c) Between 65kgs and 71 kgs

Sol: Given $\mu=68$ $\sigma=3$

ⓐ $P(z>72)$

If $x=72$

$$z=\frac{x-\mu}{\sigma}$$

$$=\frac{72-68}{3}$$

$=1.33$ $=z_1>0$

$P(z>72) = 0.5 - A(z_1)$

$= 0.5 - A(1.33)$

$= 0.5 - 0.4082$

$= 0.0918$

The no. of students above 72 kgs $= 0.0918 \times 300$

$= 27.54$

$= 27$

ⓑ $P(z \leq 64)$

If $x = 64$

$$z = \frac{x-\mu}{\sigma}$$

$$= \frac{64-68}{3}$$

$= -1.33$

$= z_1 < 0$

$P(z \leq 64) = 0.5 - A(z_1)$

$= 0.5 - A(-1.33)$

$= 0.5 - 0.4082$

$= 0.0918$

The no. of students below 64 kgs $= 0.0918 \times 300$

$= 27.54$

$= 27$

(c) $P(65 \le Z \le 71)$

If $x = 65$

$$Z = \frac{x-\mu}{\sigma}$$

$$= \frac{65-68}{3}$$

$$= -1$$

$$= Z_1$$

If $x = 71$

$$Z = \frac{x-\mu}{\sigma}$$

$$= \frac{71-68}{3}$$

$$= 1$$

$$= Z_2$$

Z_1 and Z_2 are of opposite signs

$$P(65 \le Z \le 71) = A(Z_2) + A(Z_1)$$

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

$$= 0.3413 + 0.3413$$

$$= 0.6826$$

The no. of students between 65kgs and 71 kgs

$= 0.6826 \times 300$

$= 204.78$

$= 205$

*****

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

Sol: Given, 7% of items are under 35

i.e $P(X < 35) = 7\%$

$$= \frac{7}{100}$$

$$= 0.07$$

89% of items are under 63

$$P(X < 63) = 89\%$$

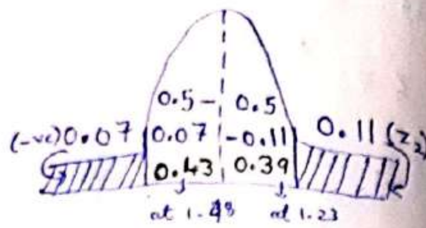
$$= \frac{89}{100} = 0.89$$

$P(x<63) = 0.89$

$P(x>63) = 1 - P(x<63)$

$= 1 - 0.89$

$P(x>63) = 0.11$



If $x = 35$

$$Z = \frac{x-\mu}{\sigma}$$

$$= \frac{35-\mu}{\sigma} = -Z_1 \quad \text{—①}$$

If $x = 63$

$$Z = \frac{x-\mu}{\sigma}$$

$$= \frac{63-\mu}{\sigma}$$

$$\frac{63-\mu}{\sigma} = Z_2 \quad \text{—②}$$

From normal distribution tables,

$Z_1 = 1.48$

$Z_2 = 1.23$

Substitute Z_1, Z_2 in ①, ②

$$\frac{35-\mu}{\sigma} = -1.48 \quad \text{—③} \Rightarrow 35 - \mu = -1.48\sigma$$

$$\frac{63-\mu}{\sigma} = 1.23 \quad \text{—④}$$

$$\frac{③}{④} \Rightarrow \frac{35-\mu}{63-\mu} = \frac{-1.48}{1.23}$$

$$35 - \mu = -1.203(63-\mu)$$

$$35 - \mu = -75.80 + 1.20\mu$$

$$2.20\mu = 110.80$$

$$\mu = 50.36$$

Mean $\mu = 50.3$

$$③ \Rightarrow \frac{35-50.3}{\sigma} = -1.48$$

$$\Rightarrow +15.3 = +1.48\sigma$$

$$\Rightarrow \sigma = 10.33$$

$$\text{Variance } \sigma^2 = (10.33)^2$$

$$= 106.70$$

$$\therefore \mu = 50, \sigma = 10$$

8) In a normal distribution, 31% of items are under 45 and 8% are over 64. Find the mean & variance of the distribution.

Sol: Given, 31% of items are under 45

$$P(X<45) = 31\%$$

$$= \frac{31}{100}$$

$$= 0.31$$

8% of items are over 64

Substitute z_1, z_2 in ①, ②

$$\frac{45-\mu}{\sigma} = -0.5 \Rightarrow 45-\mu = -0.5\sigma \quad \text{—③}$$

$$\frac{64-\mu}{\sigma} = 1.41 \Rightarrow 64-\mu = 1.41\sigma \quad \text{—④}$$

$$③-④ \Rightarrow 45-\cancel{\mu}-64+\cancel{\mu} = -0.5\sigma - 1.41\sigma$$

$$\cancel{-}19 = \cancel{+}1.91\sigma$$

$$\sigma = 9.9$$

$$\therefore \sigma = 10$$

$$\therefore \text{Variance } \sigma^2 = 100$$

$$③ \Rightarrow 45-\mu = -0.5(10) \Rightarrow 45-\mu = -5$$

$$\Rightarrow \therefore \mu = 50$$

$\boxed{\text{H.W}}$

⑨ A sales tax officer has reported that the average sales of 500 business he has to deal during a year is ₹ 36000 and S.D = 10000, find

a) The no. of business as the sales of which are ₹ 40000

b) The percentage of business of the sales are likely to range between ₹ 30000 and ₹ 40000

(10) 1000 students have written an examination. The mean of the test is 35 and S.D is 5, assuming the distribution to be normal, find

a) How many students marks lie between 25 & 40

b) How many students get more than 40

c) How many students get below 20

d)

Sol:

$\mu = 36000$ S.D $\sigma = 10000$

a) If $X = 40000$

$$Z = \frac{x - \mu}{\sigma}$$

$$= \frac{40000 - 36000}{10000}$$

$= 4000$

(b) If $X = 30000$

$z = \frac{x-\mu}{\sigma}$

$= \frac{30000-36000}{10000}$

$= \frac{-6000}{10000}$

$= -0.6$

$= z_1$

If $x = 40000$

$z = \frac{x-\mu}{\sigma}$

$= \frac{40000-36000}{10000}$

$= \frac{4000}{10000}$

$= 0.4$

$= z_2$

z_1 and z_2 are of opposite signs

$P(30000 < X < 40000) = A(z_2) + A(z_1)$

$= A(0.4) + A(-0.6)$

$= A(0.4) + A(0.6)$

$= 0.1554 + 0.2258$

$= 0.3812$

The percentage of business $= 0.3812 \times 100$

$= 38.12\%$

(10)

Sol: Given $\mu = 35$, $\sigma = 5$

(a) If $x = 25$

$z = \frac{x-\mu}{\sigma}$

$= \frac{25-35}{5}$

$= -2$

$= z_1$

If $x = 40$

$z = \frac{40-35}{5}$

$= 1$

$= z_2$

z_1 and z_2 are of opposite signs

$P(25 < X < 40) = A(z_2) + A(z_1)$

$= A(1) + A(-2)$

$= 0.3413 + 0.4772$

$= 0.8185$

$\therefore$ No. of students between 25 and 40 is $= 1000 \times 0.8185$

$= 818.5$

$= 819$

(b) If $x = 40$

$z = \frac{x - \mu}{\sigma}$

$= \frac{40 - 35}{5}$

$= 1 > 0$

$P(X > 40) = 0.5 - A(z_1)$

$= 0.5 - A(1)$

$= 0.5 - 0.3413$

$= 0.1587$

$\therefore$ No. of students greater than 40 $= 1000 \times 0.1587$

$= 158.7$

$= 159$

(c) If $x = 20$

$z = \frac{x - \mu}{\sigma}$

$= \frac{20 - 35}{5}$

$= -3 < 0$

$P(X<20) = 0.5 - A(z_1)$

$= 0.5 - A(-3)$

$= 0.5 - A(3)$

$= 0.5 - 0.4987$

$= 0.0013$

$\therefore$ No. of students below 20 is $= 0.0013 \times 1000$

$= 1.3$

$= 1$

27/10/21

Normal Approximation to Binomial Distribution:

The normal distribution can be used to approximate the B.D. Suppose the no. of success x ranges from x_1 to x_2. Then the probability of getting x_1 to x_2 success is given by

$$\sum_{r=x_1}^{x_2} {}^nC_x \, p^x q^{n-x}$$

* If n is large the calculation of binomial Probabilities is very difficult. In such cases, the binomial curve can be replaced by normal curve and the required probability is calculated by using the following formula

For any success x, the real class interval is

$\left(x-\frac{1}{2}, x+\frac{1}{2}\right)$

Hence Z_1 corresponds to lower limit of x_1 and Z_2 corresponds upper limit of x_2 i.e

$$Z_1 = \frac{\left(x_1-\frac{1}{2}\right)-\mu}{\sigma}$$

$$Z_2 = \frac{\left(x_2+\frac{1}{2}\right)-\mu}{\sigma}$$

where μ, σ are mean, S.D of Binomial Distribution

Mean $\mu = np$

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

Eg: ① Find the probability that out of 100 patients between 84 and 95 will survive a heart operation given that the chance of survival is 0.9

Sol:- Given $p=0.9$ $\quad n=100$

$q = 1-0.9$
$= 0.1$

By Normal Approximation to Binomial Distribution, we have

$$Z_1 = \frac{\left(x_1-\frac{1}{2}\right)-\mu}{\sigma}$$

$$Z_2 = \frac{\left(x_2+\frac{1}{2}\right)-\mu}{\sigma}$$

Here $x_1 = 84$ and $x_2 = 95$

Mean $\mu = np$

$= 100(0.9)$
$= 90$

Variance $\sigma^2 = npq$

$= 100(0.9)(0.1)$

$= 9$

S.D $\sigma = \sqrt{9} = 3$

$$z_1 = \frac{(x_1 - \frac{1}{2}) - \mu}{\sigma} = \frac{(84 - \frac{1}{2}) - 90}{3} = \frac{-6 - \frac{1}{2}}{3}$$

$$= \frac{-13}{6}$$

$$z_1 = \cancel{-4.33}\ -2.16$$

$$z_2 = \frac{(x_2 + \frac{1}{2}) - \mu}{\sigma} = \frac{95 + \frac{1}{2} - 90}{3} = \frac{5 + \frac{1}{2}}{3}$$

$$= \frac{11}{6}$$

$$z_2 = 1.83$$

z_1 and z_2 are of opposite signs

$P(z_1 \le z \le z_2) = A(z_2) + A(z_1)$

$= A(1.83) + A(-2.16)$

$= A(1.83) + A(2.16)$

$= 0.4664 + 0.4846$

$= 0.951$

② Find the probability that a student can correctl[y] answer 25 to 30 questions in a MCQ quiz [c]onsisting of 80 questions. Assume that in each [q]uestion with 4 choices, only one choice is correct.

Sol: Given $n=80$

$p=\frac{1}{4}$

$q=1-\frac{1}{4}$

$=\frac{3}{4}$

Mean $\mu=np$

$=80\left(\frac{1}{4}\right)$

$=20$

$\sigma^2=npq$

$=\cancel{80}^{20}\times\frac{1}{\cancel{4}_1}\times\frac{3}{\cancel{4}_1}$

$=15$

$S.D=\sqrt{15}=3.87$

$$z_1=\frac{(x_1-\frac{1}{2})-\mu}{\sigma}=\frac{25-\frac{1}{2}-20}{3.87}=\frac{5-\frac{1}{2}}{3.87}$$

$$=\frac{9}{(3.87)2}$$

$$=\cancel{2.32}\ 1.16$$

$$z_2=\frac{(x_2+\frac{1}{2})-\mu}{\sigma}=\frac{30+\frac{1}{2}-20}{3.87}=\frac{10+\frac{1}{2}}{3.87}$$

$$=\frac{21}{7.74}$$

$$=2.71$$

z_1 and z_2 are of same signs

$$P(z_1\le z\le z_2)=|A(z_2)-A(z_1)|$$

$$=|A(2.71)-A(1.16)|$$

$$=|0.4966-0.3770|$$

$$=0.1196$$

Normal Distribution:

(11)

IF X is a normal variable, find the area

a) to the left of $z = -1.78$

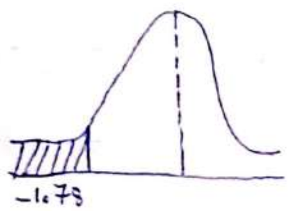
b) to the right of $z = -1.45$

c) corresponding to $-0.8 \le z \le 1.53$

d) to the left of $z = -2.52$ and to the right of $z = 1.83$

Sol:

a) left of $z = -1.78$



Required area $= 0.5 - A(\text{from } 0 \text{ to } -1.78)$

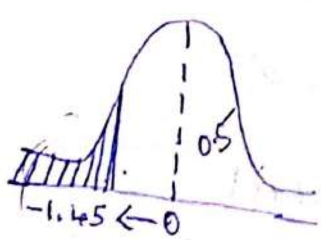
$= 0.5 - A(-1.78)$

$= 0.5 - A(1.78)$

$= 0.5 - 0.4625$

$= 0.0375$

b) to the right of $z = -1.45$



A(0 to -1.45).

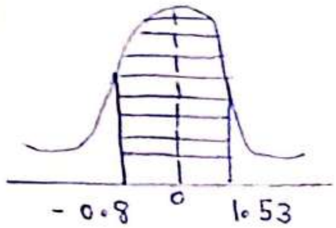
Required area $= 0.5 + A(-1.45)$

$= 0.5 + A(1.45)$

$= 0.5 + 0.4265$

$= 0.9265$

c) corresponding to $-0.8 \leq z \leq 1.53$



Required area =

If $x = -0.8$

$z = \frac{x-\mu}{\sigma}$

$= \frac{-0.8-\mu}{\sigma} = z_1$ — ①

If $x = 1.53$

$z = \frac{x-\mu}{\sigma}$

$= \frac{1.53-\mu}{\sigma} = z_2$ — ②

Required area = A(from 0 to -0.8) + A(from 0 to 1.53)

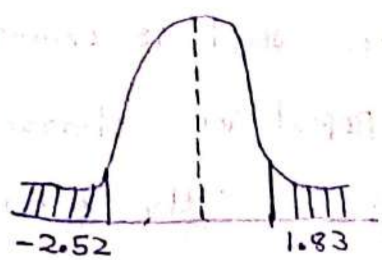
$= A(-0.8) + A(1.53)$

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

$= 0.2881 + 0.4370$

$= 0.7251$

d) left of $z = -2.52$ and right of $z = 1.83$



Required area $= 0.5 - A(0 \text{ to } -2.52) + 0.5 - A(0 \text{ to } 1.83$

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

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

$= [0.5 - 0.4941] + [0.5 - 0.4664]$

$= 0.0395$

Sampling Distributions:

Population: Population or Universe is the aggregate or totality of statistical data forming the subject of investigation.

For example, ① the population of the heights of Indians
② The population of nationalised banks in India

Definition: The no. of observations in the population is defined to be the size of the population. It may be finite or infinite. Size of the population is denoted by "N"

* Most of the times, study of entire population may not be possible to carry out and hence a part is selected from the given population.

* A portion of the population which is examined with a view to determine the population characteristic is called a Sample. The size of the sample is denoted by "n".

01/11/21

Sampling: 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 by taking into the hand from the bag and then decide to purchase it or not. Here the bag of rice is considered as population and the portion taken from the bag is considered as Sample.

Classification of Samples: Samples are classified into two ways

a) Large Sample → If the size of the sample $n > 30$, the sample is said to be large sample.

b) Small Sample → If the size of the sample $n \leq 30$, the sample is said to be small sample or simple sample.

Parameter: Any statistical measure computed from the population data is known as Parameter.

Eg: The population mean μ, Variance σ^2 are known as Parameters.

Statistic: Any Statistical measure computed from Sample data is known as Statistic.

Eg: Sample Mean $\bar{x}$, Sample Variance s^2 are known as Statistic

Formulae:

① The correction factor is

$$C.F = \frac{N-n}{N-1}$$

* There are two types of samplings

a) With replacement b) Without replacement

* The total no. of samples with replacement is given by N^n.

* The total no. of samples without replacement is given by ${}^N C_n$.

*** SAQ

Eg: ① What is the value of correction factor if $n=5$ $N=200$

Sol: Given size of the population $N=200$

Size of the sample $n=5$

$$\text{Correction Factor} = \frac{N-n}{N-1}$$

$$= \frac{200-5}{200-1}$$

$$= \frac{195}{199}$$

$$= 0.9799$$

KKKKKK B.

(2) A population consist of 5 numbers 2, 3, 6, 8 and 11. Consider all possible samples of size 2 which can be drawn i) With replacement ii) Without replacement from the population. Find

i) The mean of the population

ii) The Standard Deviation of the population

iii) The mean of the sampling distribution of mean

iv) The S.D of the sampling distribution of mean

Sol: Given Size of the population $N = 5$

Size of the sample $n=2$

i) Mean of Population $\mu = \frac{\sum_{i=1}^{n} x_i}{n}$

$$= \frac{2+3+6+8+11}{5}$$

$$= \frac{30}{5}$$

$$\mu = 6$$

ii) Standard Deviation of Population

Variance $\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n}$

$$= \frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5}$$

$$= \frac{16+9+0+4+25}{5}$$

$$= \frac{54}{5}$$

$$\sigma^2 = 10.8$$

$S.D = \sqrt{10.8}$

$= 3.286$

iii) <u>With replacement:</u>

The total no.of samples with replacement is given by

$N^n = 5^2$

$= 25$ samples

The samples are

$$\left\{\begin{matrix} (2,2) & (2,3) & (2,6) & (2,8) & (2,11) \\ (3,2) & (3,3) & (3,6) & (3,8) & (3,11) \\ (6,2) & (6,3) & (6,6) & (6,8) & (6,11) \\ (8,2) & (8,3) & (8,6) & (8,8) & (8,11) \\ (11,2) & (11,3) & (11,6) & (11,8) & (11,11) \end{matrix}\right\}$$

Find the mean of each sample in the above set

$$\left\{\begin{matrix} 2 & 2.5 & 4 & 5 & 6.5 \\ 2.5 & 3 & 4.5 & 5.5 & 7 \\ 4 & 4.5 & 6 & 7 & 8.5 \\ 5 & 5.5 & 7 & 8 & 9.5 \\ 6.5 & 7 & 8.5 & 9.5 & 11 \end{matrix}\right\} \text{ — (I)}$$

Mean of Sampling Distribution of Mean

$$\mu_{\bar{x}} = \frac{(2+2.5+4+5+6.5)+(2.5+3+4.5+5.5+7)+(4+4.5+6+7+8.5)+(5+5.5+7+8+9.5)+(6.5+7+8.5+9.5+11)}{25}$$

$$\mu_{\bar{x}} = \frac{150}{25} = 6$$

iv) Variance of Sampling distribution of mean:

$$\sigma_{\bar{x}}^2 = \frac{\sum_i (x_i - \bar{x})^2}{n}$$

$$= \frac{\cancel{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2}}{25}$$

$$= \cancel{16 + 9 + 0 + 4 +}$$

$$= \frac{(2-6)^2 + (2.5-6)^2 + (4-6)^2 + (6.5-6)^2 + (5-6)^2 + (2.5-6)^2 + (3-6)^2 + (4.5-6)^2 + (5.5-6)^2 + (7-6)^2 + (4-6)^2 + (4.5-6)^2 + (6-6)^2 + (7-6)^2 + (8.5-6)^2 + (5-6)^2 + (5.5-6)^2 + (7-6)^2 + (8-6)^2 + (9.5-6)^2 + (6.5-6)^2 + (7-6)^2 + (8.5-6)^2 + (9.5-6)^2 + (11-6)^2}{25}$$

$$\sigma_{\bar{x}}^2 = 5.40$$

S.D $\sigma_{\bar{x}} = \sqrt{5.40}$

$= 2.32$

Without Replacement:

The total no.of samples without replacement is given by $^5C_2 = \frac{5 \times 4}{2} = 10$

The 10 samples are

$$\left\{\begin{matrix} (2,3) & (2,6) & (2,8) & (2,11) \\ & (3,6) & (3,8) & (3,11) \\ & & (6,8) & (6,11) \\ & & & (8,11) \end{matrix}\right\}$$

Find the mean of above sample

$$\begin{Bmatrix} 2.5 & 4 & 5 & 6.5 \\ & 4.5 & 5.5 & 7 \\ & & 7 & 8.5 \\ & & & 9.5 \end{Bmatrix}$$

$$\mu_{\bar{x}} = \frac{2.5+4+5+6.5+4.5+5.5+7+7+8.5+9.5}{10}$$

$$= \frac{60}{10}$$

$$= 6$$

Variance of Sampling Distribution of Mean

$$\sigma_{\bar{x}}^2 = \Sigma \frac{(x_i - \bar{x})^2}{N}$$

$$= \frac{(2.5-6)^2 + (4-6)^2 + (5-6)^2 + (6.5-6)^2 + (4.5-6)^2 + (5.5-6)^2 + (7-6)^2 + (7-6)^2 + (8.5-6)^2 + (9.5-6)^2}{10}$$

$$= \frac{40.5}{10}$$

$$\sigma_{\bar{x}}^2 = 4.05$$

$$S.D = \sqrt{4.05}$$

$$\doteq 2.012$$

③ A Population consist of 5, 10, 14, 18, 13, 24 Consider all samples of size 2 which can be drawn without replacement. Find

i) The mean of the population

ii) The S.D of the Population

iii) The mean of the sampling distribution of mean

iv) The S.D of the sampling distribution of mean

Sol: Given Size of the population $N=6$

Size of the sample $n=2$

i) Mean of the population $= \frac{\sum_{i=1}^{n} x_i}{N}$

$$= \frac{5+10+14+18+13+24}{6}$$

$$\mu = \frac{84}{6} = 14$$

ii) Variance $\sigma^2 = \frac{\sum(x_i-\bar{x})^2}{n}$

$$= \frac{(5-14)^2+(10-14)^2+(14-14)^2+(18-14)^2+(13-14)^2+(24-14)^2}{6}$$

$$= \frac{81+16+0+16+1+100}{6} = \frac{214}{6}$$

$$\sigma^2 = 35.6$$

S.D $= \sqrt{35.6}$

$= 5.96$

iii) The total number of samples without replacement is given by ${}^6C_2 = \frac{6\times5}{2} = 15$

$$\left\{\begin{matrix} (5,10) & (5,14) & (5,18) & (5,13) & (5,24) \\ & (10,14) & (10,18) & (10,13) & (10,24) \\ & & (14,18) & (14,13) & (14,24) \\ & & & (18,13) & (18,24) \\ & & & & (13,24) \end{matrix}\right\} \text{— (I)}$$

Find the mean of above sample

7.5	9.5	11.5	9	14.5
	12	14	11.5	17
		16	13.5	19
		[redacted]	15.5	21
				18.5

Mean of the sampling distribution of mean

$$\mu_{\bar{x}} = \frac{7.5+9.5+11.5+9+14.5+12+14+11.5+17+16+13.5+19+15.5+21+18.5}{15}$$

$$= \frac{210}{15}$$

$$\mu_{\bar{x}} = 14$$

iv) Variance of Sampling distribution of mean

$$\sigma_{\bar{x}}^2 = \sum \frac{(x_i-\bar{x})^2}{N}$$

$$= \frac{(7.5-14)^2+(9.5-14)^2+(11.5-14)^2+(9-14)^2+(14.5-14)^2+(12-14)^2+(14-14)^2+(11.5-14)^2+(17-14)^2+(16-14)^2+(13.5-14)^2+(19-14)^2+(15.5-14)^2+(21-14)^2+(18.5-14)^2}{15}$$

$$= \frac{214}{15}$$

$$= 14.266$$

Standard Deviation $\sigma_{\bar{x}} = \sqrt{14.26} = 3.78$

(4) Samples of size 2 are taken from the population 3, 6, 9, 15, 27 with replacement. Find

i) The mean of the population

ii) The S.D of the population

iii) The mean of the sampling distribution of mean

iv) The S.D of the sampling distribution of mean

Sol: Given size of population $N=5$

Size of sample $n=2$

i) Mean of the population $= \dfrac{\sum_{i=1}^{n} x_i}{N}$

$$= \frac{3+6+9+15+27}{5}$$

$$= \frac{60}{5}$$

$$\mu = 12$$

ii) Standard Deviation of the population $\sigma^2 = \dfrac{\sum (x_i - \bar{x})^2}{N}$

$$= \frac{(3-12)^2+(6-12)^2+(9-12)^2+(15-12)^2+(27-12)^2}{5}$$

$$= \frac{81+36+9+9+225}{5}$$

$$= \frac{360}{5} = 72$$

$$S.D = \sqrt{72}$$

$$= 8.485$$

iii) The total number of samples with replacement is

$= 5^2 = 25$

The samples are

$$\begin{Bmatrix} (3,3) & (3,6) & (3,9) & (3,15) & (3,27) \\ (6,3) & (6,6) & (6,9) & (6,15) & (6,27) \\ (9,3) & (9,6) & (9,9) & (9,15) & (9,27) \\ (15,3) & (15,6) & (15,9) & (15,15) & (15,27) \\ (27,3) & (27,6) & (27,9) & (27,15) & (27,27) \end{Bmatrix} \quad \text{(I)}$$

Mean of above sample

$$\begin{Bmatrix} 3 & 4.5 & 6 & 9 & 15 \\ 4.5 & 6 & 7.5 & 10.5 & 16.5 \\ 6 & 7.5 & 9 & 12 & 18 \\ 9 & 10.5 & 12 & 15 & 21 \\ 15 & 16.5 & 18 & 21 & 27 \end{Bmatrix}$$

Mean of sampling distribution of mean

$$\mu_{\bar{x}} = \frac{3+4.5+6+9+15+4.5+6+7.5+10.5+16.5+6+7.5+9+12+18+9+10.5+12+15+21+15+16.5+18+21+27}{25}$$

$$= \frac{300}{25}$$

$$\mu_{\bar{x}} = 12$$

iv) Variance of the sampling distribution of mean

$$\sigma_{\bar{x}}^2 = \frac{\begin{bmatrix}(3-12)^2+(4.5-12)^2+(6-12)^2+(9-12)^2+(15-12)^2+(4.5-12)^2+(6-12)^2+(7.5-12)^2+ \\ (10.5-12)^2+(16.5-12)^2+(6-12)^2+(7.5-12)^2+(9-12)^2+(12-12)^2+(18-12)^2+ \\ (9-12)^2+(10.5-12)^2+(12-12)^2+(15-12)^2+(21-12)^2+(15-12)^2+(16.5-12)^2+ \\ (18-12)^2+(21-12)^2+(27-12)^2\end{bmatrix}}{25}$$

$\sigma_{\bar{x}}^2 = \frac{900}{25}$

$= 36$

S.D $\sigma_{\bar{x}} = \sqrt{36}$

$= 6$

(5) If the population is 3,6,9,15,27.

i) List all possible samples of size 3 that can be taken without replacement from the given population.

Find

a) The mean of the population

b) S.D of the population

c) Mean of the sampling distribution of mean

d) S.D of the sampling distribution of mean

Sol: Size of the population $N=5$

Size of the sample $n=3$

a) Mean $\mu = \frac{\sum_{i=1}^{n} x_i}{n}$

$= \frac{3+6+9+15+27}{5}$

$= \frac{60}{5}$

$\mu = 12$

b) S.D of the population

$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$

$= \frac{(3-12)^2 + (6-12)^2 + (9-12)^2 + (15-12)^2 + (27-12)^2}{5}$

$= \frac{81+36+9+9+225}{5} = \frac{360}{5}$

$\Rightarrow \sigma^2 = 72$

S.D $\sigma = \sqrt{72}$

$= 8.4853$

Without Replacement

iii) c)
The total no. of samples without replacement is given by $^{N}C_{n} = {}^{5}C_{3}$

$= 10$ samples

The list of samples are

$$\left\{ \begin{array}{ccccc} (3,6,9) & (3,6,15) & (3,9,15) & (3,6,27) & (3,9,27) \\ (3,15,27) & (6,9,15) & (6,9,27) & (6,15,27) & (9,15,27) \end{array} \right\}$$

Find the mean of each sample

$$\left\{ \begin{array}{ccccc} 6 & 8 & 9 & 12 & 13 \\ 15 & 10 & 14 & 16 & 17 \end{array} \right\}$$

Mean of sampling distribution of mean

$$\mu_{\bar{x}} = \frac{6+8+9+12+13+15+10+14+16+17}{10}$$

$$= \frac{120}{10}$$

$$\mu_{\bar{x}} = 12$$

d) S.D of the sampling distribution of mean

$$\sigma^2 = \frac{\sum(x_i - \mu_{\bar{x}})^2}{n}$$

$$= \frac{(6-12)^2+(8-12)^2+(9-12)^2+(12-12)^2+(13-12)^2+(15-12)^2+(10-12)^2+(14-12)^2+(16-12)^2+(17-12)^2}{10}$$

$$= \frac{36+16+9+1+9+4+4+16+25}{10} = \frac{120}{10}$$

$\sigma_{\bar{x}}^2 = 12$

$\sigma_{\bar{x}} = \sqrt{12} = 3.464$

NOTE: If $\bar{x}$ be the mean of a sample of size "n" drawn from a population with mean μ and S.D σ then

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Standard Error: It is used for calculating the difference between the expected value and the observed value.

The standard error of mean is given by $S.E = \frac{\sigma}{\sqrt{n}}$

① The variance of a population is 2. The size of the sample collected from the population is 169. What is the standard error of mean.

Sol: Given

Size of the sample $n = 169$

Variance $\sigma^2 = 2 \Rightarrow$ S.D $\sigma = \sqrt{2} = 1.414$

S.E of mean $= \frac{\sigma}{\sqrt{n}}$

$= \frac{1.414}{\sqrt{169}}$

$= \frac{1.414}{13}$

$= 0.108$

② A random sample of size 100 is taken from an infinite population having the mean $\mu = 76$ and Variance $\sigma^2 = 256$. What is the probability that $\bar{x}$ will be between 75 and 78

Sol: Given, Size of sample $n = 100$

Mean $\mu = 76$

Variance $\sigma^2 = 256 \Rightarrow$ S.D $\sigma = \sqrt{256} = 16$

If $\bar{x} = 75$ $\quad z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{75 - 76}{16/\sqrt{100}} = \dfrac{-1}{1.6} = -0.625 = z_1$

If $\bar{x} = 78$ $\quad z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{78 - 76}{16/\sqrt{100}} = \dfrac{2}{1.6} = 1.25 = z_2$

$$P(75 \le \bar{x} \le 78) = A(z_2) + A(z_1)$$
$$= A(1.25) + A(-0.625)$$
$$= A(1.25) + A(0.625)$$
$$= 0.3944 + 0.2324$$
$$= 0.6268$$

③ A random sample of size 64 is taken from a normal Population with $\mu = 51.4$ & $\sigma = 6.8$. What is the probability that the mean of the sample will be

a) exceeds 52.9

b) falls between 50.5 & 52.3

c) less than 50.6

Sol: Given, Size of the sample $n = 64$

Mean $\mu = 51.4$

S.D $\sigma = 6.8$

i) $P(z > 52.9)$

If $x = 52.9$

$$Z = \frac{x-\mu}{\sigma/\sqrt{n}} = \frac{52.9 - 51.4}{6.8/\sqrt{64}} = \frac{1.5}{6.8/8} = \frac{1.5}{0.85} = 1.76 = z_1 > 0$$

$$P(z > 52.9) = 0.5 - A(1.76)$$
$$= 0.5 - 0.4608$$
$$= 0.0392$$

ii) $P(50.5 < x < 52.3)$

If $x = 50.5$

$$z = \frac{x-\mu}{\sigma/\sqrt{n}}$$
$$= \frac{50.5 - 51.4}{6.8/\sqrt{64}}$$
$$= \frac{-0.9}{0.85}$$
$$= -1.05 = z_1$$

If $x = 52.3$

$$z = \frac{x-\mu}{\sigma/\sqrt{n}}$$
$$= \frac{52.3 - 51.4}{6.8/\sqrt{64}}$$
$$= \frac{0.9}{0.85}$$
$$= 1.05$$
$$= z_2$$

z_1 and z_2 are of opposite sign

$$P(z_1 < z < z_2) = P(50.5 < z < 52.3) = A(z_2) + A(z_1)$$
$$= A(1.05) + A(-1.05)$$
$$= A(1.05) + A(1.05)$$
$$= 2(0.3531)$$
$$= 0.7062$$

iii) $P(z < 50.6)$

If $x = 50.6$

$$z = \frac{x - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{50.6 - 51.4}{6.8/\sqrt{64}}$$

$$= \frac{-0.8}{0.85}$$

$$= -0.94 = z_1 < 0$$

$$P(z<50.6) = 0.5 - A(z_1)$$
$$= 0.5 - A(-0.94)$$
$$= 0.5 - A(0.94)$$
$$= 0.5 - 0.3264$$
$$= 0.1736$$

Chebyshev's Inequality: Chebyshev's inequality helps to derive bounds on probabilities when only the mean or both the mean and variance of the probability distribution are given.

Chebyshev's Theorem: If a probability distribution ha mean μ and S.D σ, the probability of getting a value which deviates μ by atleast K_σ is atmost $\frac{1}{K^2}$ i.e

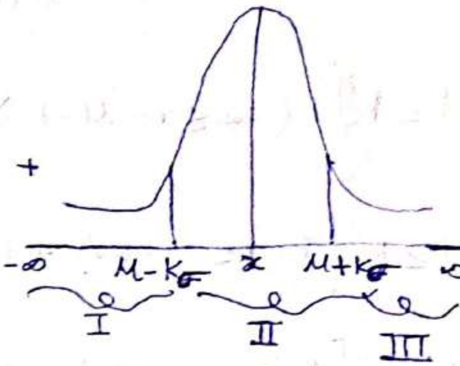
$P[|x-\mu| \geq K\sigma] \leq \frac{1}{K^2}$ is the probability associated ith set of outcomes for which the value of a ndom variable having the given probability stribution is such that $|x-\mu| \geq K\sigma$

Proof: Given Mean $E(x) = \mu$ & Variance $V(x) = \sigma^2$

By definition of variance, we have

$$\sigma^2 = V(x) = E\left[x - E(x)\right]^2$$
$$= E(x-\mu)^2$$
$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 . f(x)\, dx$$

Now dividing the interval into 3 parts



From the above, we have

$$\sigma^2 = \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)dx$$

Neglecting the second integral, since second integral is always a positive value.

$$\sigma^2 \geq \underbrace{\int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\, dx}_{\mathrm{I}} + \underbrace{\int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)dx}_{\mathrm{II}} \quad \text{—— (1)}$$

From First Integral	From Second Integral
$x \leq \mu - k\sigma$	$x \geq \mu + k\sigma$
$x - \mu \leq -k\sigma$	$x - \mu \geq k\sigma$
$-(x-\mu) \geq k\sigma$	$(x-\mu)^2 \geq k^2\sigma^2$ —— (3)
$(x-\mu)^2 \geq k^2\sigma^2$ —— (2)	

Substitute (2) & (3) in (1)

$$\sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 f(x)\,dx + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 f(x)\,dx$$

$$\sigma^2 \geq k^2\sigma^2 \left[\int_{-\infty}^{\mu-k\sigma} f(x)\,dx + \int_{\mu+k\sigma}^{\infty} f(x)\,dx \right]$$

$$1 \geq k^2 \left[\int_{-\infty}^{\mu-k\sigma} f(x)\,dx + \int_{\mu+k\sigma}^{\infty} f(x)\,dx \right]$$

$$1 \geq k^2 \left\{ \left[P(-\infty \leq x \leq \mu - k\sigma) \right] + P\left[(\mu + k\sigma \leq x \leq \infty) \right] \right\}$$

$$1 \geq k^2 \left[P(x \leq \mu - k\sigma) + P(x \geq \mu + k\sigma) \right]$$

$$1 \geq k^2 \left[P(x - \mu \leq -k\sigma) + P(x - \mu \geq k\sigma) \right]$$

$$1 \geq k^2 \left[P(-(x-\mu) \geq k\sigma) + P(x-\mu) \geq k\sigma \right]$$

$$1 \geq k^2 \left[P\,|x - \mu| \geq k\sigma \right]$$

$$\frac{1}{k^2} \geq \left[P|x-\mu| \geq k\sigma \right]$$

$$P\left[|x-\mu| \geq k\sigma \right] \leq \frac{1}{k^2}$$

UNIT - III

Estimation and Test of Hypothesis

Introduction:

A statistical inference, we use the sample and apply a suitable statistical method for drawing conclusions about the unknown properties of the population so that, we obtain the answer to a problem.

There are two types of problems under statistical inference

1) Estimation
2) Testing of Hypothesis

Estimate: An estimate is a statement made to find an unknown population parameter.

Estimator: The procedure (or) rule to determine an unknown population parameter is called an Estimator.

Eg: Sample mean is an estimator of Population mean.

Estimation: To use the statistics, obtained from the samples as an estimate of the unknown Parameter of the population from which the sample is drawn.

There are two types of estimation

1) Point estimation
2) Interval estimation

* Estimation means approximating a quantity to the required accuracy.

Point Estimation: If an estimate of the population parameter is given by a single value, then the estimate is called a point estimation of the parameter.

Eg: If the height of a student is measured as 162cm then the measurement gives a point estimation.

* A point estimate of a parameter (θ) is a single numerical value which is computed from a given sample and acts as an approximation of the unknown exact value of the parameter.

* A point estimator is denoted by '$\hat{\theta}$'.

Interval Estimation: An interval estimate of a population parameter (θ) is an interval of the form $\hat{\theta}_L < \theta < \hat{\theta}_u$ where $\hat{\theta}_L$ and $\hat{\theta}_u$ depend on the value of the statistic $\hat{\theta}$ for a particular sample and also on the sampling distribution of $\hat{\theta}$.

* If an estimate of a population parameter is given by two different values between which the parameter may be considered to lie, then the estimate is called an interval estimation of the parameter.

Eg: If the height of the student is given as (163 ± 3.5)cm Then the height lies between 159.5 cms and 166.5 cms and the measurement gives an interval estimation.

Unbiased Estimator: A point estimator $(\hat{\theta})$ is said to be unbiased estimator of the parameter (θ) if $\boxed{E(\hat{\theta}) = \theta}$

In other words, $\boxed{E(\text{Statistic}) = \text{parameter}}$ then the statistics is said to be unbiased estimator of the parameter.

Biased Estimator: A point estimator $(\hat{\theta})$ is said to be biased estimator if $\boxed{E(\hat{\theta}) \neq \theta}$ i.e

$\boxed{E(\text{Statistic}) \neq \text{Parameter}}$

Parameter $\rightarrow$ Population $\rightarrow \mu, \sigma^2$

Statistic $\rightarrow$ Sample $\rightarrow \bar{x}, s^2$

Properties of estimators:

* A good estimator is one which is closed to the true value of the parameter.

* The important properties are

1) Consistency
2) Unbiasedness
3) Efficiency
4) Sufficiency

Theorem: The sample mean $(\bar{x})$ is an unbiased estimator of the population mean (μ)

Proof: Let $x_1, x_2, \ldots x_n$ be a random sample drawn from a given population with mean μ

To prove that $E(\bar{x}) = \mu$

Consider, $E(\bar{x}) = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right]$

$$= E.\frac{1}{n}\left[\sum_{i=1}^{n} x_i\right]$$

$$= \frac{1}{n} E\left[x_1 + x_2 + \cdots + x_n\right]$$

$$= \frac{1}{n}\left[E(x_1) + E(x_2) + \cdots + E(x_n)\right]$$

$$= \frac{1}{n}\left[\mu + \mu + \mu + \cdots + \mu\right]$$

$$= \frac{1}{\cancel{n}}(\cancel{n}\mu)$$

$$E(\bar{x}) = \mu$$

Hence Proved

15/11/21

Example: The mean of a random sample is an unbiased estimate of the mean of the population 3, 6, 9, 15, 27

a) List all possible samples of size 3 that can be taken without replacement from the finite population

) Calculate the mean of each sample & assign h sample a probability of $\frac{1}{10}$. Verify that the n of the sample $\bar{x} = 12$, which is equal to the an of the population, θ i.e $E(\bar{x}) = \theta$

Prove that $\bar{x}$ is an unbiased estimate of θ.

Sol: Given $N=5$ and $n=3$

a) The total no. of samples without replacement is given by $^{N}C_{n} = {^5C_3} = 10$ samples

$$\left\{ \begin{array}{lll} (3,6,9) & (3,6,15) & (3,6,27) \\ (6,9,15) & (6,9,27) & (9,15,27) \\ (3,9,15) & (3,9,27) & (6,15,27) \\ (3,15,27) & & \end{array} \right\} \text{— (1)}$$

b) To prove that $\bar{x}$ is an unbiased estimate of Population θ i.e. $E(\bar{x}) = \theta$

The population mean $\theta = \dfrac{3+6+9+15+27}{5}$

$$= \frac{60}{5}$$

$$\theta = 12$$

Find mean of each sample of (1)

$$\left\{ \begin{array}{lll} 6 & 8 & 12 \\ 10 & 14 & 17 \\ 9 & 13 & 16 \\ 15 & & \end{array} \right\}$$

Now assign each sample a probability of $\frac{1}{10}$ i.e

x_i	6	8	12	10	14	17	9	13	16	15
$P(x_i)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

To find,

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

$$= 6\left(\frac{1}{10}\right) + 8\left(\frac{1}{10}\right) + 12\left(\frac{1}{10}\right) + 14\left(\frac{1}{10}\right) + 10\left(\frac{1}{10}\right) + 17\left(\frac{1}{10}\right) + 9\left(\frac{1}{10}\right) + 13\left(\frac{1}{10}\right) + 16\left(\frac{1}{10}\right) + 15\left(\frac{1}{10}\right)$$

$$= \frac{120}{10}$$

$$= 12$$

$\therefore E(\bar{x}) = \theta$

$\therefore \bar{x}$ is an unbiased estimate of population parameter θ

(2) Suppose that we observe a random variable having the binomial distribution

a) Show that $\frac{x}{n}$ is an unbiased estimate of the binomial parameter p.

b) Show that $\frac{x+1}{n+2}$ is not an unbiased estimate of binomial parameter p.

Sol:

a) To show that, $E\left(\frac{x}{n}\right) = p$

Consider $E\left(\frac{x}{n}\right) = \frac{1}{n} E(x)$

$$= \frac{1}{n}[\mu]$$

[$\because$ In B.D $\mu = np$]

$$= \frac{1}{\cancel{n}}(\cancel{n}p)$$

$\therefore E\left(\frac{x}{n}\right) = p$

$\therefore \frac{x}{n}$ is an unbiased estimate of p.

b) To show that, $E\left(\frac{x+1}{n+2}\right) \neq p$

Consider $E\left(\frac{x+1}{n+2}\right) = \frac{1}{n+2} E(x+1)$

$$= \frac{E(x)}{n+2} + \frac{E(1)}{n+2}$$

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

$$E\left(\frac{x+1}{n+2}\right) = \frac{np+1}{n+2} \neq p$$

$\therefore \frac{x+1}{n+2}$ is not an unbiased estimate of p.

Formulae:

* To find the size of sample

$\rightarrow n = \left(\frac{Z_{\alpha/2}}{E}\right)^2 PQ$ [$\because$ If proportion is given]

$\rightarrow n = \left(\frac{Z_{\alpha/2}\,\sigma}{E}\right)^2$ [$\because$ If σ is given]

The max error is $E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

$\rightarrow$ The confidence interval for mean is

$$\left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Confidence Limits:

* For 95% confidence limit is 1.96
* For 99% confidence limit is 2.58
* For 90% confidence limit is 1.645
* For 98% confidence limit is 2.33

③ What is the maximum error one can expect to make with probability 0.90 when using the mean of a random sample of size $n = 64$ to estimate the mean of the population with $\sigma^2 = 2.56$

Sol: Given $n = 64, \sigma^2 = 2.56 \Rightarrow \sigma = 1.6$

$Z_{\alpha/2} = 0.90 = 90\%$

$Z_{\alpha/2} = 1.645$

Maximum error $E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

$= 1.645 \times \frac{1.6}{\sqrt{64}}$

$= \frac{2.632}{8}$

$= 0.329$

Confidence Interval :- A confidence interval displays the prob that a parameter will fall between a pair of values around the mean. Confidence interval is used to measure uncertainty in a sample variable. The confidence level represents the frequency of acceptable confidence intervals that contain the true value of unknown parameter

④ Assuming that $\sigma=20$, how large a random sample is taken to assert with probability 0.95 that the sample will not differ from the true mean by more than three points error.

Sol: Given $\sigma=20$, $E=3$, $Z_{\alpha/2}=95\%$

$=1.96$

Size of the sample, $n=\left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2$

$=\left(\frac{1.96\times 20}{3}\right)^2$

$=(13.06)^2$

$=170.737$

$\cong 171$

⑤ The dean of a college wants to use the mean of a random sample to estimate the average amount of time students to get from one class to next class and he want to be able to assert with 99% confidence that the error is atmost 0.25 min. If it can be preresumed from experience that $\sigma=1.40$ min. How large a sample will he have to take.

Sol: Given $E=0.25$ min

$\sigma=1.40$ min

$Z_{\alpha/2}=99\%$

$=2.58$

Size of Sample $n = \left(\frac{Z_{\alpha/2}\,\sigma}{E}\right)^2$

$= \left(\frac{2.58 \times 1.40}{0.25}\right)^2$

$= (14.448)^2$

$= 208.74$

$\cong 209$

⑥ What is the size of the sample required to estimate an unknown proportion with a maximum error of 0.06 with atleast 95% confidence.

Sol: Given $E = 0.06$

Given, proportion i.e $P = \frac{1}{2}$

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

$Z_{\alpha/2} = 95\% = 1.96$

Size of the sample $n = \left(\frac{Z_{\alpha/2}}{E}\right)^2 Pq$

$= \left(\frac{1.96}{0.06}\right)^2 \frac{1}{2} \times \frac{1}{2}$

$= \frac{1067.11}{4}$

$= 266.77$

$\cong 267$

16/11/21

(7) The mean and S.D of a population are 11.795 and 14.054. What can one assert with 95% confidence about the maximum error if $\bar{x} = 11.795$ and $n = 50$. Construct 95% confidence interval for the true mean.

Sol: Given $\mu = 11.795$, $\sigma = 14.054$, $\bar{x} = 11.795$ $n = 50$

The confidence interval for mean is

$$\left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

$$= \left[11.795 - \frac{1.96(14.054)}{\sqrt{50}}, 11.795 + \frac{1.96(14.054)}{\sqrt{50}}\right]$$

$$= \left[11.795 - 3.895, 11.795 + 3.895\right]$$

$$= \left[7.89, 15.69\right]$$

$$= (7.89, 15.69)$$

(8) A random sample of size 81 was taken with variance 20.25 and mean is 32, construct 98% confidence interval.

Sol: Given $\bar{x} = 32$, $n = 81$, $\sigma^2 = 20.25 \Rightarrow \sigma = 4.5$

$Z_{\alpha/2} = 2.33$

Confidence interval is $\left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$

$$= \left[32 - \frac{2.33(4.5)}{\sqrt{81}}, 32 + \frac{2.33(4.5)}{\sqrt{81}}\right]$$

$$= \left[32 - 1.165,\ 32 + 1.165\right]$$

$$= \left[30.835, 33.165\right]$$

(9) A random sample of 500 points on a heated plate result in an avg. temperature of 73.54 degrees with a S.D of 2.79 degrees. Find a 99% confidence interval for the average temperature of the plate.

Sol: $\bar{x} = 73.54,\ \sigma = 2.79,\ Z_{\alpha/2} = 2.58,\ n = 500$

Confidence interval is $\left(\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

$$= \left(73.54 - \frac{(2.79)(2.58)}{\sqrt{500}},\ 73.54 + \frac{(2.58)(2.79)}{\sqrt{500}}\right)$$

$$= (73.54 - 0.32,\ 73.54 + 0.32)$$

$$= (73.22, 73.86)$$

Estimation: To use the statistic obtained by the samples as an estimate to predict the unknown parameter of the population from which the sample is drawn.

Applications of Normal Distribution:

(1) Data obtained from Psychological, Physical and Biological measurements approximately follow Normal Distribution. Eg: Scores, Heights & Weights

(2) Most of the distributions like Binomial, Poisson can be approximated to Normal Distribution.

- If n is very large and neither p nor q is very small, Binomial Distribution tends to Normal Distribution.
- If the parameter $\lambda \to \infty$, then Poisson Distribution tends to Normal Distribution.

(3) Normal Distribution is applicable to many applied Problems in Kinetic Theory of Gases & Fluctuations in magnitude of an electric current.

(4) For large samples, any statistic (sample mean, sample S.D etc) approximately follows Normal Distribution.

(5) Normal curve is used to find confidence limits of population parameters.

(6) Normal distribution finds large application in Statistical Quality Control in industry for finding control limits.

Conditional prob :- Happening of one Event after happening of other event. Here we use 2 events. In general prob. we have only one Event

Always first event will be in denominator.
" " second Event will be in numerator

i.e A first
B-second means $P(B|A)$

Eg:- I take the class after going to college.
B event — A - Event

$P(B/A)$ A event happen first then B happens

Total theorem of prob:- If $B_1, B_2, \ldots, B_n$ are the set of Exhaustive & mutually Exclusive Events in Sample space S & A is any Event which is subset of $\bigcup_{i=1}^{n} B_i$ then

$$P(A) = \sum_{i=1}^{n} P(B_i) \, P(A/B_i)$$

Proof :- If $B_1, B_2 \ldots B_n$ are set of Exhaustive & mutually Exclusive Events in 'S'

$$\therefore S = \bigcup_{i=1}^{n} B_i$$

The inner circle represents an Event A which can occur along with $B_1, B_2, \ldots, B_n$.

i.e $A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n)$

Here $(A \cap B_1), (A \cap B_2) \ldots (A \cap B_n)$ are mutually Exclusive.

$$P(A) = P[(A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n)]$$
$$= P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n)$$
$$= P(B_1) . P(A/B_1) + P(B_2) . P(A/B_2) + \ldots + P(B_n) \, P(A/B_n)$$

$$P(A) = \sum_{i=1}^{n} P(B_i) \, P(A/B_i).$$

CSE-D

11/10/22 · Presentees no.

R → $K_3, K_8, L_0, M_2, M_5, N_4$

$N_5, N_9, P_0, P_2, P_5, P_7,$

$P_9, Q_2, Q_4, Q_7, R_2,$

L.E → Sunny Joel.

→ Seema kousal.

12/10/22 Presentees.

$K_3, L_2, L_8, L_9, M_5, M_7, M_8, M_9,$
$N_6, N_8, N_9, P_0, P_5, P_7, P_9, Q_2,$
$Q_4, Q_6, Q_8, R_2, R_3, R_4$

L.E → 21, 22, 25, 24.

13/10/22 Presentees.

$K_3, K_8, L_0, L_8, M_7, M_8, M_9, N_4, N_8,$
$N_9, P_0, P_5, P_9, Q_2, Q_4, Q_6, R_2, R_3, R_4,$
$R_7, M_2, M_6, N_5, Q_7, M_3, L_2, Q_5, Q_8$

L.E → 22, 25, 24, 20, M_5.

18/10/22 K, L, M, N, P, Q, R.

$L_0, L_1, L_4, L_6, L_7, L_8,$ $N_2, N_3,$ $Q_1,$ Q_8
$Q_3, Q_9, R_2,$

L.E — ~~20, 21, 22~~ ~~24~~ 23

19/10/22

$K_9, L_0, L_4, L_9, M_4, M_6, N_0, N_1, N_3, N_9$
$P_2, P_3, P_9, Q_2, Q_4, Q_8.$

L.E → 23

20/10/22 Ab → $K_4, L_0, L_3, L_4, L_5, L_7,$
$M_0, M_1, N_3, P_1, P_2, P_3, P_9, Q_1, Q_4, Q_8,$

L.E - Pes — 20, 21, 23, 24, 25, Kashi Vishwanath.

CSM-A

11/10/22 Present no

6, 7, 8, 10, 12, 13, 14, 15, 19, 20,
21, 22, 23, 24, 25, 27, 29, 31, 32, 33,
35, 40, 41, 45, 47, 48, 50, 51, 52,
53, 57, 58, 59, 61, L.E - all absent

13/10/22 Present no.

2, 4, 6, 7, 8, 9, 11, 12, 13, 15, 16, 17,
18, 19, 20, 21, 22, 23, 25, 27, 31, 32,
33, 34, 35, 36, 39, 40, 41, 44, 45,
48, 49, 50, 51, 52, 53, 57, 59, 61,
62,

L.E — 3,

15/10/22 Present no.

2, 4, 5, 7, 8, 13, 15, 21, 25, 30, 32, 35,
36, 39, 42, 44, 51, 52, 54, 58, 59, 62,
64, 61.

L.E — 5 6.

20/10/22 — absentees.

8, 16, 17, 18, 23, 27, 34, 41, 43, 62,

L.E : 2, 6,

22/10/22 Presentees

1, 6, 7, 8, 12, 16, 21, 25, 22, 26,
31, 33, 35, 39, 41, 45, 46, 50, 51, 61

L.E → 1, 4, 6, 5,

25/10/22 Presentees

1, 4, 10, 25, 32, 33, 35, 36, 40,
51, 63.

L.E — 2, 4, 5,

[illegible]

15/11/21

Tuesday

UNIT-III

ESTIMATION & TEST OF HYPOTHESIS

Test of Hypothesis: A statistical hypothesis is a statement about the parameter of one or more population.

* Testing of hypothesis is a process for deciding whether to accept or reject the hypothesis.
* There are 2 types of hypothesis:

a) Null Hypothesis

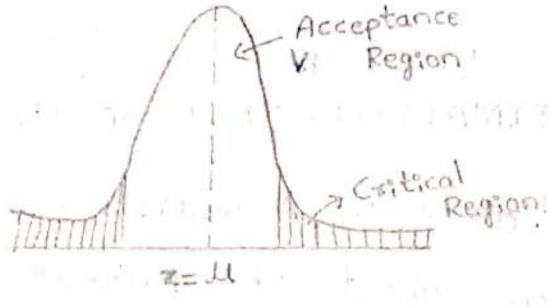
b) Alternative Hypothesis

Null Hypothesis: A hypothesis of no difference is called Null Hypothesis. It is denoted by H_0.

Alternative Hypothesis: A hypothesis which contradict the null hypothesis is called Alternative hypothesis It is denoted by H_1.

Critical Region: A region corresponding to a statistic in the sample space which leads to reject null hypothesis is called Critical region.

* The region which leads to accept null hypothesis is called Acceptance region.
* The value of test statistic which separates rejection region & acceptance region is called Critical Value (or) Significant Value.



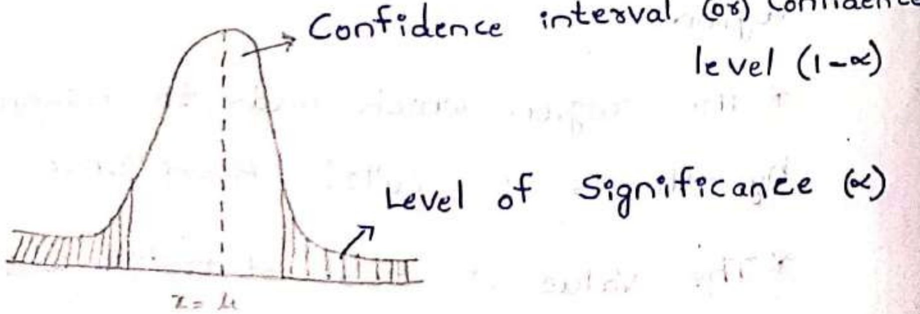
17/11/21

Level of Significance [LOS]:

* The probability with which we will reject a null hypothesis when it is true.
* It is denoted by α.
* LOS is the cut-off value which determines whether the researcher will reject null hypothesis or accept null hypothesis.

Confidence Interval:

* The probability with which we will accept a null hypothesis when it is true.
* It is denoted by $1-\alpha$.
* The probability to accept a true null hypothesis is denoted by $1-\alpha$



LOS (α)		Confidence level ($1-\alpha$)	
0.05	5%	0.95	95%
0.01	1%	0.99	99%
0.1	10%	0.90	90%

***

Errors of Sampling: There are two types of errors -

a) Type I error

b) Type II error

Type I error: It is the error of rejecting null hypothesis H_0 when it is true. The probability of making a Type I error is denoted by α i.e. Level of Significance.

Type II error: It is the error of accepting the null hypothesis H_0 when it is false. The Type II error is denoted by β i.e.

$$P(\text{reject } H_0, \text{when it is true}) = P(\text{type-I error}) = \alpha$$

$$P(\text{accept } H_0, \text{when it is false}) = P(\text{type-II error}) = \beta$$

***

Two-tailed Test: Two-tailed test is a test of hypothesis in which alternative hypothesis is expressed by using the symbol not equal to ($\neq$).

* In two-tailed test, the rejection will be on both the sides i.e on the right and left side.

* In this test of hypothesis, the critical region is divided into two parts i.e

The critical region under the right tail =

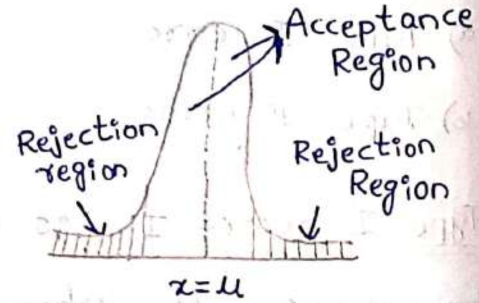
= critical area under the left tail

= Half of the total area

= $\frac{1}{2}$(Probability of Rejection)

= $\frac{1}{2}(\alpha)$

= $\frac{\alpha}{2}$ (Two-tailed)



Two-tailed test

***

One-tailed Test:

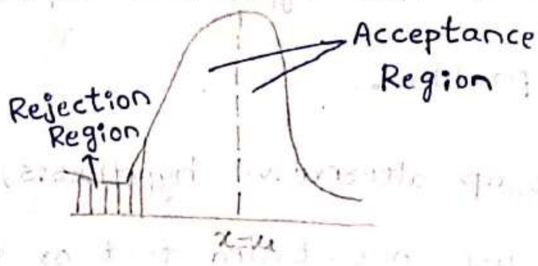
* One-tailed Test is a test of hypothesis in which the alternative hypothesis is expressed by using the symbol less than or greater than.

* In this test of hypothesis, the entire critical region lies in the one tail of the distribution.

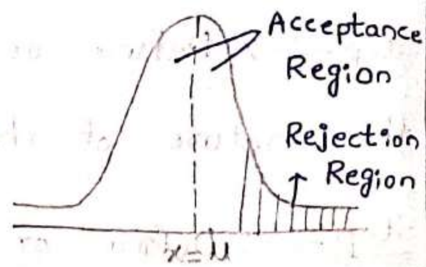
* Critical region will lie at the right tail if alternative hypothesis has symbol greater than (>).

* Critical region will lie at the left tail if alternative hypothesis has symbol less than (<).

* In one-tailed test, the rejection will be only on one side i.e either left or right.



Left tailed test



Right tailed test

***

NOTE:

1) If the alternative hypothesis is two-tailed test, it can be written as $H_1: \mu \neq \mu_0$

2) If the alternative hypothesis is right-tailed test, it can be written as $H_1: \mu > \mu_0$

3) If the alternative hypothesis is left ~~to~~ tailed test, it can be written as $H_1: \mu < \mu_0$

***

Critical Values of Z_α:

	Level of Significance		
	1% (0.01)	5% (0.05)	10% (0.1)
Two-tailed test	$Z_\alpha = 2.58$	$Z_\alpha = 1.96$	$Z_\alpha = 1.645$
Right-tailed test	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
Left-tailed test	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$	$Z_\alpha = -1.28$

SAQ ***

Procedure for Testing of Hypothesis:

Step-1 → Define or set up null hypothesis depending on the nature of the problem.

Step-2 → Define or set up alternative hypothesis, in order to decide whether to use one-tailed test or two-tailed test i.e

$H_1: \mu \neq \mu_0$ [Two tailed]

$H_1: \mu > \mu_0$ [Right tailed]

$H_1: \mu < \mu_0$ [Left tailed]

Step-3 → Write the LOS α Value from the given Problem. If α is not given by default take $\alpha = 5\%$ level of significance i.e 0.05

Step-4 → Write the test Statistic and calculate z value i.e $Z_{\text{calculated value}}$

Step-5 → Now compare $Z_{\text{calculated value}}$ and $Z_{\text{tabulated value}}$

a) If $Z_{\text{calculated}} < Z_{\text{tabulated}}$ then we accept null hypothesis

b) If $Z_{\text{calculated}} > Z_{\text{tabulated}}$ then we reject the null hypothesis

Methods of Testing of Hypothesis:

1 Test of Significance for single mean

2 Test of Significance for difference of two means

3 Test of Significance for single proportion

4 Test of significance for difference of two proportions

20/11/21

1 Test of Significance for single mean:

Step-1 → Set up null hypothesis depending on the nature of the problem i.e $H_0: \mu = \mu_0$

Step-2 → Set up alternative hypothesis i.e

$H_1: \mu \neq \mu_0$ [Two-tailed]
$H_1: \mu > \mu_0$ [Right tailed]
$H_1: \mu < \mu_0$ [Left tailed]

Step-3 → Write α value from the given problem, if not take $\alpha = 0.05$ or 5% LOS

Step-4 → Calculate the test statistic.

The test statistic for single mean is

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Step-5 → Compare $Z_{calculated}$ & $Z_{tabulated}$ values

→ If $Z_{calculated} < Z_{tabulated}$ then we ~~reject~~ accept null hypothesis.

(SAQ) ***

Procedure for Testing of Hypothesis:

Step-1 → Define or set up null hypothesis depending on the nature of the problem.

Step-2 → Define or set up alternative hypothesis, in order to decide whether to use one-tailed test or two-tailed test i.e

$H_1: \mu \neq \mu_0$ [Two tailed]

$H_1: \mu > \mu_0$ [Right tailed]

$H_1: \mu < \mu_0$ [Left tailed]

Step-3 → Write the LOS α Value from the given problem. If α is not given by default take $\alpha = 5\%$ level of significance i.e 0.05

Step-4 → Write the test Statistic and calculate z value i.e $Z_{\text{calculated value}}$

Step-5 → Now compare $Z_{\text{calculated value}}$ and $Z_{\text{tabulated value}}$

a) If $Z_{\text{calculated}} < Z_{\text{tabulated}}$ then we accept null hypothesis

b) If $Z_{\text{calculated}} > Z_{\text{tabulated}}$ then we reject the null hypothesis

Methods of Testing of Hypothesis:

1. Test of Significance for single mean
2. Test of Significance for difference of two means
3. Test of Significance for single proportion
4. Test of significance for difference of two proportions

20/11/21

1. Test of Significance for Single mean:

Step-1 → Set up null hypothesis depending on the nature of the problem i.e $H_0 : \mu = \mu_0$

Step-2 → Set up alternative hypothesis i.e

$H_1 : \mu \neq \mu_0$ [Two-tailed]

$H_1 : \mu > \mu_0$ [Right tailed]

$H_1 : \mu < \mu_0$ [Left tailed]

Step-3 → Write α value from the given problem, if not take $\alpha = 0.05$ or 5% LOS

Step-4 → Calculate the test statistic.

The test statistic for single mean is

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Step-5 → Compare $Z_{calculated}$ & $Z_{tabulated}$ values

→ If $Z_{calculated} < Z_{tabulated}$ then we ~~reject~~ accept null hypothesis.

If $Z_{calculated} > Z_{tabulated}$ then we reject null hypothesis

① The mean lifetime of a sample of 100 lights produced by a company is found to be 1560 hrs with a population S.D of 90 hrs. Test the hypothesis for $\alpha = 0.05$ that the mean life time of the lights produced by the company is 1580 hrs.

Sol: Given sample size $n = 100$

Mean of the Population $\mu = 1580$ hrs

Mean of the sample $\bar{x} = 1560$ hrs

S.D $\sigma = 90$ hrs

Step-1 → Set up null hypothesis i.e $H_0 : \mu = 1580$ hrs

The mean life time of the lights produced by the company is 1580 hrs.

Step-2 → Set up alternative hypothesis i.e $H_1 : \mu \neq 1580$ hrs

The mean life time of the lights produced by the company is not 1580 hrs. [Two-tailed Test]

Step-3 → Given $\alpha = 0.05$

From the table, $Z_{tabulated} = 1.96$

Step-4 → Calculate the test statistic, i.e

$$Z_{calculated} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{1560 - 1580}{90/\sqrt{100}}$$

$$= \frac{-20}{\frac{90}{10}}$$

$$= -\frac{20}{9}$$

$|Z_{calculated}| = |-2.22| = 2.22$

Step-5 → Compare $Z_{calculated}$ & $Z_{tabulated}$ Values

$Z_{cal} > Z_{tab}$

Hence, we reject null hypothesis ($\because Z_{cal} > Z_{tab}$)

i.e We accept alternative hypothesis.

Hence, the mean life time of the lights produced by the company is not 1580 hrs.

② In a random sample of 60 workers, the average time taken by them to get into work is 33.8 min with a S.D of 6.1 min. Can we reject the null hypothesis $\mu = 32.6$ in favor of alternative hypothesis $\mu > 32.6$ min at $\alpha = 0.05$ LOS.

Sol: Given $n = 60$

$\mu = 32.6$ min

$\bar{x} = 33.8$ min

$\sigma = 6.1$ min

Step-1 → $H_0 : \mu = 32.6$

Step-2 → $H_1 : \mu > 32.6$ [Right tailed]

Step-3 → $\alpha = 0.05$

From table, $Z_{tabulated} = 1.645$

Step-4 →

$$Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{33.8 - 32.6}{6.9/\sqrt{60}}$$

$$= \frac{1.2}{0.787}$$

$Z_{calculated} = 1.52$

Step-5 → $Z_{cal} = 1.52$, $Z_{tab} = 1.645$

$Z_{tabulated} > Z_{calculated}$ ie $Z_{calculated} < Z_{tabulated}$

We accept null hypothesis

③ An ambulance service claims that it takes on the average less than 10 min to reach its destination in emergency calls. A sample of 36 calls has a mean of 11 min and the variance 16 min. Test the claim at 0.05 LOS.

Sol: Given $n = 36$

$\mu = 10$

$\bar{x} = 11$

$\sigma^2 = 16 \Rightarrow \sigma = 4$

Step-1 → $H_0: \mu = 10$

Step-2 → $H_1: \mu < 10$ [Left tailed]

Step-3 → $\alpha = 0.05$

From table, $Z_{tabulated} = -1.645$

$|Z_{tabulated}| = 1.645$

Step-4 →

$$Z_{cal} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

$$= \frac{11-10}{\frac{4}{\sqrt{36}}}$$

$$= \frac{1}{4/6}$$

$$= \frac{6}{4}$$

$$Z_{cal} = 1.5$$

Step-5 → $Z_{cal} < Z_{tab}$

We ~~reject~~ accept null hypothesis

(4) 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 cakulate 95% confidence interval for the population.

Sol: Given $n = 400$

$\bar{x} = 40$

$\mu = 38$

$\sigma = 10$

Step-1 → $H_0 : \mu = 38$

Step-2 → $H_1 : \mu \neq 38$ [Two tailed Test]

Step-3 → $\alpha = 0.05$

From table, $Z_{tabulated} = 1.96$

Step-4 → $$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

$$= \frac{40-38}{10/\sqrt{400}}$$

$$= \frac{2}{\frac{10}{20}}$$

$Z_{calculated} = 4$

<u>Step-5</u> → $Z_{calculated} > Z_{tabulated}$

Hence, we reject null hypothesis

95% confidence interval for mean is

$$\left(\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

$$= \left(40 - 1.96\left(\frac{10}{\sqrt{400}}\right), 40 + 1.96\left(\frac{10}{\sqrt{400}}\right)\right)$$

$$= \left(40 - \frac{1.96}{2}, 40 + \frac{1.96}{2}\right)$$

$$= (39.02, 40.98)$$

(5) A sample of 64 students have a mean weight 70kgs. Can this be regarded as a sample from a population with mean 56 kgs and S.D 25 kgs.

<u>Sol:</u> Given $n = 64$

$\mu = 56$

$\bar{x} = 70$

$\sigma = 25$

<u>Step-1</u> → $H_0: \mu = 56$

<u>Step-2</u> → $H_1: \mu \neq 56$ [Two-tailed Test]

<u>Step-3</u> → $\alpha = 0.05$

From table, $Z_{tabulated} = 1.96$

Step-4 →
$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$
$$= \frac{70 - 56}{25/\sqrt{64}}$$
$$= \frac{14}{25/8}$$
$$Z_{calculated} = 4.48$$

Step-5 → $Z_{calculated} > Z_{tabulated}$

∴ We reject null hypothesis

$\boxed{22/11/21}$

[2] Test of Significance for difference of two means:

Procedure:

Step-1 → Set up null hypothesis i.e $H_0 : \mu_1 = \mu_2$

Step-2 → Set up alternative hypothesis

$H_1 : \mu_1 \neq \mu_2$ [Two-tailed Test]

$H_1 : \mu_1 > \mu_2$ [Right tailed Test]

$H_2 : \mu_1 < \mu_2$ [Left tailed Test]

Step-3 → Write LOS α value from the given problem if not take $\alpha = 0.05$

Step-4 → Calculate the test statistic

The test statistic for difference of two means is

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \left[\because \sigma_1^2 \neq \sigma_2^2\right]$$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \qquad \left[\sigma_1^2 = \sigma_2^2\right]$$

Step-5 → Calculate $Z_{calculated}$ & $Z_{tabulated}$ Values

a) If $Z_{calculated} < Z_{tabulated}$, we accept null hypothesis

b) If $Z_{calculated} > Z_{tabulated}$, we reject null hypothesis

① The research investigator is interested in studying whether there is a significant difference in the salaries of MBA in 2 metropolitan cities. A random sample of size 100 from Mumbai earns an average income of Rs. 20,150. Another random sample of size 60 from Chennai earns an average income of Rs 20,250. If the Variance of both the populations are given as σ_1^2 = Rs 40,000 & σ_2^2 = Rs 32,400 respectively.

Sol: Given $n_1 = 100$, $n_2 = 60$

$\bar{x}_1 = 20,150$

$\bar{x}_2 = 20,250$

$\sigma_1^2 = 40,000$

$\sigma_2^2 = 32,400$

Step-1 → Set up null hypothesis i.e H_0: There is no significant difference between two cities i.e

$$H_0: \mu_1 = \mu_2$$

Step-2 → Set up alternative hypothesis i.e H_1: There is significant difference between two cities i.e

$$H_1: \mu_1 \neq \mu_2 \text{ [Two-tailed Test]}$$

Step-3 → $\alpha = 0.05$

$$Z_{tabulated} = 1.96$$

Step-4 → Calculate the test statistic i.e

$$Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{20,150 - 20,250}{\sqrt{\frac{40000}{100} + \frac{\cancel{32400}^{540}}{\cancel{60}_{1}}}}$$

$$= \frac{-100}{\sqrt{400+540}}$$

$$= \frac{-100}{\sqrt{940}}$$

$$|Z_{cal}| = |-3.26| = 3.26$$

Step-5 → Compare $Z_{calculated}$ & $Z_{tabulated}$ Values

$$Z_{cal} > Z_{tab}$$

∴ We reject null hypothesis

② The mean of 2 large samples of sizes 1000 & 2000 members are 67.5 inches & 68 inches are respectively. Can the samples be regarded as drawn from the same population with S.D 2.5 inches.

<u>Sol:</u> Given, $n_1 = 1000$

$n_2 = 2000$

$\bar{x}_1 = 67.5$

$\bar{x}_2 = 68$

$\sigma = \sigma_1 = \sigma_2 = 2.5$

<u>Step-1</u> → Set up null hypothesis

$H_0 : \mu_1 = \mu_2$

<u>Step-2</u> →

$H_1 : \mu_1 \neq \mu_2$ [Two-tailed Test]

<u>Step-3</u> → $\alpha = 0.05$

From distribution table, $Z_{tabulated} = 1.96$

<u>Step-4</u> →

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{67.5 - 68}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}}$$

$$= \frac{-0.5}{2.5\,(0.0387)}$$

$$= \frac{-0.5}{0.09675} = -5.16$$

$Z_{calculated} = |-5.16| = 5.16$

Step-5 → $Z_{cal} > Z_{tab}$

We reject null hypothesis

③ Samples of students are drawn from 2 universities and their weights are in kilograms, mean & S.D are calculated and shown below. Make a large sample test to test the significance of difference between the means.

	Mean	S.D	Size
University A	55	10	400
University B	57	15	100

Sol: Given $n_1 = 400$

$n_2 = 100$

$\bar{x}_1 = 55$

$\bar{x}_2 = 57$

$\sigma_1 = 10 \Rightarrow \sigma_1^2 = 100$

$\sigma_2 = 15 \Rightarrow \sigma_2^2 = 225$

Step-1 → $H_0: \mu_1 = \mu_2$

Step-2 → $H_1: \mu_1 \neq \mu_2$ [Two-tailed Test]

Step-3 → $\alpha = 0.05$

$Z_{tabulated} = 1.96$ (From table)

Step-4 → $$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{55-57}{\sqrt{\frac{100}{400}+\frac{225}{100}}}$$

$$= \frac{-2}{\sqrt{2.5}}$$

$$= \frac{-2}{1.58}$$

$$Z_{cal} = |-1.26| = 1.26$$

Step-5 → $Z_{cal} < Z_{tab}$

∴ We accept null hypothesis

(4) The average marks scored by 32 boys is 72 with S.D of 8. While, for 36 girls is 70 with a S.D of 6. Does this indicate that the boys perform better than girls at LOS 0.05.

Sol: Given $n_1 = 32$

$n_2 = 36$

$\bar{x}_1 = 72$

$\bar{x}_2 = 70$

$\sigma_1 = 8 \Rightarrow \sigma_1^2 = 64$

$\sigma_2 = 6 \Rightarrow \sigma_2^2 = 36$

Step-1 → Set up null hypothesis $H_0: \mu_1 = \mu_2$

Step-2 → Set up alternative hypothesis

$H_1: \mu_1 > \mu_2$ [Right Tailed Test]

Step-3 → $\alpha = 0.05$

$Z_{tabulated} = 1.645$

Step-4 → $Z_{cal} = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

$$= \frac{72-70}{\sqrt{\frac{64}{32} + \frac{36}{36}}}$$

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

$$= \frac{2}{\sqrt{3}}$$

$$Z_{cal} = 1.154$$

Step-5 → $Z_{cal} < Z_{tabulated}$

∴ We accept null hypothesis

∴ We conclude that there is no significant difference in marks scored by the boys and girls.

⑤ A sample of height of 6400 Englishmen has a mean of 67.85 inches & a S.D of 2.56 inches while a sample of height of 1600 Australians has a mean of 68.55 inches with a S.D of 2.52 inches. Do the data indicate that the Australians are on the average taller than the Englishmen at LOS 0.01

Sol: $n_1 = 6400$ $\quad \bar{x}_1 = 67.85$ $\quad \sigma_1 = 2.56$

$n_2 = 1600$ $\quad \bar{x}_2 = 68.55$ $\quad \sigma_2 = 2.52$

Step-1 → Null Hypothesis

$$H_0 : \mu_1 = \mu_2$$

Step-2 → $H_1: \mu_1 < \mu_2$ [Left-tailed Test]

Step-3 → $\alpha = 0.01$

$$Z_{tabulated} = |-2.33| = |2.33|$$

Step-4 →

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{67.85 - 68.55}{\sqrt{\frac{(2.56)^2}{6400} + \frac{(2.52)^2}{1600}}}$$

$$= \frac{-0.7}{\sqrt{4.993 \times 10^{-3}}}$$

$$= |-9.906|$$

$$Z_{calculated} = 9.90$$

Step-5 → $Z_{calculated} > Z_{tabulated}$

We reject null hypothesis

⑥ The mean height of 50 main students who participated in sports is 68.2 inches with a S.D of 2.5. The mean height of 50 main students who have not participated n the sports is 67.2 inches with a S.D of 2.8. est the hypothesis that the height of the students ho participated in sports is more than the students who havenot participated in the sports.

Sol: Given $n_1 = 50$

$n_2 = 50$

$\bar{x}_1 = 68.2$

$\bar{x}_2 = 67.2$

$\sigma_1 = 2.5$ and $\sigma_2 = 2.8$

Step-1→ Set up Null Hypothesis $H_0: \mu_1 = \mu_2$

Step-2→ Set up Alternative Hypothesis

$H_1: \mu_1 > \mu_2$ [Right tailed test]

$\alpha = 0.05$

Step-3→

$Z_{tabulated} = 1.645$

Step-4→ Calculate the test statistic

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{68.2 - 67.2}{\sqrt{\frac{(2.5)^2}{50} + \frac{(2.8)^2}{50}}}$$

$$= \frac{1}{\sqrt{\frac{14.09}{50}}}$$

$$= \frac{1}{0.530}$$

$$= 1.8863$$

$Z_{calculated} = 1.8863$

Step-5→ $Z_{calculated} > Z_{tabulated}$

...le reject null hypothesis

23/11/21

Method 3: Test of Significance for Single Proportion

Step-1 → Set up null hypothesis i.e $H_0: P = P_0$

Step-2 → Set up alternative hypothesis i.e

$H_1: P \neq P_0$ [Two-tailed Test]

$H_1: P > P_0$ [Right-tailed Test]

$H_2: P < P_0$ [Left-tailed Test]

Step-3 → Write Level of Significance, α value from the given problem, if not take $\alpha = 0.05$

Step-4 → Calculate the test statistic

The test statistic for single proportion is

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

where P = Population proportion

$Q = 1 - P$

p = Sample proportion

$p = \frac{x}{n}$ where

x = no. of favourable cases

Step-5 → Compare $Z_{calculated}$ & $Z_{tabulated}$ values

If $Z_{cal} < Z_{tab}$, we accept null hypothesis

If $Z_{cal} > Z_{tab}$, we reject null hypothesis

① In a sample of 1000 people in Karnataka, 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% LOS.

Sol: $n = 1000, P = \frac{1}{2}, Q = \frac{1}{2}$

$$P = 0.5, Q = 0.5$$

$$x = 540, p = \frac{x}{n} = \frac{540}{1000} = 0.54 \Rightarrow p = 0.54$$

Step-1 → Set up null hypothesis H_0:
Both rice and wheat eaters are equally popular in the state i.e $H_0: P = 0.5$

Step-2 → Set up alternative hypothesis H_1: Both rice and wheat eaters are not equally popular in the state.

$H_1: P \neq 0.5$ [Two-tailed Test]

Step-3 → Given $\alpha = 1\%$
$= 0.01$

$$\Rightarrow Z_{tab} = 2.58$$

Step-4 → Calculate the test statistic

$$Z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}}$$

$$= \frac{0.04}{0.0158}$$

$$= 2.529$$

$$Z_{cal} = 2.529$$

Step-5 → Compare Z_{cal} & Z_{tab} values

$$Z_{cal} < Z_{tab}$$

We accept null hypothesis

Both rice and wheat eaters are equally popular in the state.

② 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 $n = 600, x = 325$

$$P = \frac{1}{2} = 0.5, \quad Q = \frac{1}{2} = 0.5$$

$$p = \frac{x}{n} = \frac{325}{600} = 0.5417$$

Step-1 → $H_0: P = 0.5$

Step-2 → $H_1: P > 0.5$ [Right-tailed Test]

Step-3 → $\alpha = 0.05$

$$\Rightarrow Z_{tab} = 1.645$$

Step-4 → $Z_{cal} = \dfrac{p-P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0.5417-0.5}{\sqrt{\dfrac{0.5 \times 0.5}{600}}} = 2.04$

$$Z_{cal} = 2.04$$

Step-5 → $Z_{cal} > Z_{tab}$

We reject null hypothesis.

③ Experience had shown that 20% of a manufactured product is of top quality. In one day production of 400 articles only, 50 are of top quality. Test the hypothesis at 0.05 LOS.

Sol: Given $n=400, x=50$

$$p=\frac{x}{n} = \frac{50}{400} = 0.125$$

Given $P=20\% = \frac{20}{100} \Rightarrow P=0.2$

$$Q=1-P = 1-0.2 \Rightarrow Q=0.8$$

Step-1 → $H_0: P=0.2$

Step-2 → $H_1: P \neq 0.2$ [Two-tailed Test]

Step-3 → $\alpha=0.05$

$$\Rightarrow Z_{tab}=1.96$$

Step-4 → $Z_{cal} = \dfrac{p-P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0.125-0.2}{\sqrt{\dfrac{0.2\times0.8}{400}}} = -3.75$

$$\Rightarrow |Z_{cal}| = 3.75$$

Step-5 → $Z_{cal} > Z_{tab}$

We reject null hypothesis

NOTE:

Confidence interval for single proportion

$$\left(\underset{(p)}{p} - Z_{\alpha/2}\sqrt{\frac{PQ}{n}}, \underset{(small)}{p} + Z_{\alpha/2}\sqrt{\frac{PQ}{n}}\right)$$

***

④ In a random sample of 160 workers exposed to a certain amount of radiation, 24 experienced sum ill effects. Construct 99% confidence interval for the corresponding true percentage.

Sol: Given $n=160, x=24$

$$P=\frac{x}{n}=\frac{24}{160}=0.15 \Rightarrow P=0.15$$

$$q=1-P=1-0.15 \Rightarrow q=0.85$$

99% confidence interval $\Rightarrow$ 1% LOS

$= 0.01$

$$Z_{\alpha/2}=Z_{\alpha}=2.58$$

The confidence interval is $\left(P-Z_{\alpha/2}\sqrt{\frac{Pq}{n}}, P+Z_{\alpha/2}\sqrt{\frac{Pq}{n}}\right)$

$$=\left(0.15-2.58\sqrt{\frac{0.15\times0.85}{160}}, 0.15+2.58\sqrt{\frac{0.15\times0.85}{160}}\right)$$

$$=(0.07, 0.222)$$

⑤ In a random sample of 100 packages taken by flight, 13 had some damage. Construct 95% confidence interval for the true proportion of the damage.

Sol: $n=100, x=13$

$$P=\frac{x}{n}=\frac{13}{100} \Rightarrow P=0.13$$

$$q=1-P=1-0.13 \Rightarrow q=0.87$$

95% confidence interval = 5% LOS

$= 0.05$

$$\Rightarrow Z_{\alpha/2}=Z_{\alpha}=1.96$$

The confidence interval is $\left(P - Z_{\alpha/2}\sqrt{\frac{Pq}{n}}, P + Z_{\alpha/2}\sqrt{\frac{Pq}{n}}\right)$

$$= \left(0.13 - 1.96\sqrt{\frac{0.13 \times 0.87}{100}}, 0.13 + 1.96\sqrt{\frac{0.13 \times 0.87}{100}}\right)$$

$$= (0.064, 0.195)$$

⑥ In a hospital, 480 females and 520 male babies were born in a week. Do these figures confirm the hypothesis that males & females are born in equal number.

Sol: $n = 480 + 520$

$= 1000$

$x = 480$

$$p = \frac{x}{n} = \frac{480}{1000} = 0.48$$

$$P = \frac{1}{2} = 0.5, \quad Q = \frac{1}{2} = 0.5$$

Step-1 → $H_0 : P = 0.5$

Step-2 → $H_1 : P \neq 0.5$ [Two-tailed]

Step-3 → $\alpha = 0.05$

$Z_{tab} = 1.96$

Step-4 → $$Z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.48 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}}$$

$= -1.26$

$|Z_{cal}| = 1.26$

Step-5 → $Z_{cal} < Z_{tab}$

We accept null hypothesis

(7) A manufacturer claimed that atleast 95% of the equipment which he supplied to a factory are good. An examination of a sample of 200 pieces of equipment revealed that 18 are defective. Test the claim at 5% LOS.

Sol:- $n = 200$, $x = 200 - 18$
$= 182$

$p = \frac{x}{n} = \frac{182}{200} = 0.91$ $\quad P = 95\% = \frac{95}{100} = 0.95$

$Q = 1 - P = 1 - 0.95 = 0.05$

Step-1 → $H_0: P = 0.95$

Step-2 → $H_1: P \neq 0.95$ [Two-tailed Test]

Step-3 → $\alpha = 5\%$ LOS

$\Rightarrow Z_{tab} = 1.96$

Step-4 → $Z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.91 - 0.95}{\sqrt{\frac{0.95 \times 0.05}{200}}} = -2.59$

$|Z_{cal}| = 2.59$

Step-5 → $Z_{cal} > Z_{tab}$

We reject null hypothesis.

Method-4: Test of Significance for difference of two proportions

Step-1 → Set up null hypothesis

H_0: There is no significant difference between two proportions

i.e $H_0: P_1 = P_2$

Step-2 → Set up alternative hypothesis

H_1: There is significant difference between two proportions

$H_1: P_1 \neq P_2$ [Two-tailed Test]

$H_1: P_1 > P_2$ [~~Left~~ Right-tailed Test]

$H_1: P_1 < P_2$ [~~Right~~ Left tailed Test]

Step-3 → Write Level of Significance α value from the given problem if not take $\alpha = 0.05$

Step-4 → Calculate the test statistic

The test statistic for difference of two proportions is

$$Z = \frac{P_1 - P_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{where } \hat{P} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}, \ \hat{Q} = 1 - \hat{P}$$

$$P_1 = \frac{x_1}{n_1}, \ P_2 = \frac{x_2}{n_2}$$

Step-5 → Compare $Z_{calculated}$ & $Z_{tabulated}$ values

If $Z_{cal} < Z_{tab}$, we accept null hypothesis

If $Z_{cal} > Z_{tab}$, we reject null hypothesis

24/11/21

① Random samples of 400 men & 600 women are asked whether they would like to have a flyover near their residence. 200 men and 325 women are in favour of the proposal. Test the hypothesis that proportions of men and women in favour of the proporsal are same at 5% LOS.

sol: Given $n_1 = 400, n_2 = 600$

$x_1 = 200, \; x_2 = 325$

$P_1 = \frac{x_1}{n_1} = \frac{200}{400} = \frac{1}{2} = 0.5$

$P_2 = \frac{x_2}{n_2} = \frac{325}{600} = 0.5416$

Step-1→ Set up null hypothesis H_0: There is no significant difference between the option of men & women in favour of the proporsal.

$H_0': P_1 = P_2$

Step-2→ Set up alternative hypothesis H_1: There is a significant difference between men & women in favour of the proporsal

$H_1: P_1 \neq P_2$ [Two tailed Test]

Step-3→ $\alpha = 5\%$ LOS

$= 0.05$

$Z_{tab} = 1.96$

Step-4 → Calculate the test statistic

$$Z_{cal} = \frac{P_1 - P_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$\hat{P} = \frac{n_1P_1 + n_2P_2}{n_1+n_2} = \frac{400(0.5)+600(0.5416)}{400+600}$$

$$\hat{P} = 0.525$$

$$\hat{q} = 1-\hat{P} = 1-0.525$$

$$\hat{q} = 0.475$$

$$Z_{cal} = \frac{0.5 - 0.5416}{\sqrt{(0.525)(0.475)\left[\frac{1}{400}+\frac{1}{600}\right]}}$$

$$= -1.29$$

$$|Z_{cal}| = 1.29$$

Step-5 → $Z_{cal} < Z_{tab}$

We accept null hypothesis

∴ We conclude that, there is no significant difference between men & women in favour of the proposal.

② In a sample of 600 students of a certain college 400 are found to use ball pens. In another college from a sample of 900 students, 450 were found to use ball pens. Test whether two colleges are significantly different w.r.to the habit of using ball pens.

Sol: $n_1 = 600$ $n_2 = 900$

$x_1 = 400$ $x_2 = 450$

$$P_1 = \frac{x_1}{n_1} = \frac{400}{600} = 0.67$$

$$P_2 = \frac{x_2}{n_2} = \frac{450}{900} = 0.5$$

Step-1 → $H_0 : P_1 = P_2$

Step-2 → $H_1 : P_1 \neq P_2$ [Two-tailed Test]

Step-3 → $\alpha = 0.05$

$Z_{tab} = 1.96$

Step-4 →

$$\hat{P} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

$$= \frac{600(0.67) + 900(0.5)}{600 + 900}$$

$$= 0.568$$

$$\hat{q} = 1 - 0.568$$
$$= 0.432$$

$$Z_{cal} = \frac{P_1 - P_2}{\sqrt{\hat{P}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{0.67 - 0.5}{\sqrt{(0.568)(0.432)\left[\frac{1}{600} + \frac{1}{900}\right]}}$$

$$= 6.51$$

Step - 5 → $Z_{cal} > Z_{tab}$

We reject null hypothesis

③ Before an increase on excise duty on tea 500 People out of a sample of 900 found to have the habit of having tea. After an increase on excise duty, 250 have the habit of having tea among 1100. Test the hypothesis at 5% LOS.

Sol: $n_1 = 900$ $\quad n_2 = 1100$

$x_1 = 500$ $\quad x_2 = 250$

$P_1 = \frac{x_1}{n_1} = \frac{500}{900} = 0.55$

$P_2 = \frac{x_2}{n_2} = \frac{250}{1100} = 0.22$

Step - 1 → $H_0 : P_1 = P_2$

Step - 2 → $H_1 : P_1 > P_2$ [Right tailed test]

Step - 3 → $\alpha = 0.05$

$Z_{tab} = 1.645$

Step - 4 → $\hat{P} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$

$= \frac{900(0.55) + 1100(0.22)}{900 + 1100}$

$= 0.3685$

$\hat{q} = 1 - P = 1 - 0.3685$

$\hat{q} = 0.6315$

$$Z_{cal} = \frac{0.55 - 0.22}{\sqrt{(0.3685)(0.6315)\left[\frac{1}{900} + \frac{1}{1100}\right]}}$$

$$= 15.21$$

Step-5 → $Z_{cal} > Z_{tab}$

We reject null hypothesis

Q) In an investigation on the machine performance, the following results are obtained

	No. of units inspected	No. of defectives
Machine 1	375	17
Machine 2	450	22

Test whether is there any significant difference in the machines at $\alpha = 0.05$

Sol:- $n_1 = 375$ $\quad n_2 = 450$

$x_1 = 17$ $\quad x_2 = 22$

$$P_1 = \frac{x_1}{n_1} = \frac{17}{375} = 0.045$$

$$P_2 = \frac{x_2}{n_2} = \frac{22}{450} = 0.048$$

Step-1 → $H_0 : P_1 = P_2$

Step-2 → $H_1 : P_1 \neq P_2$ [Two-tailed Test]

Step-3 → $\alpha = 0.05$

$Z_{tab} = 1.96$

Step-4 →
$$\hat{p} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

$$= \frac{375(0.045) + 450(0.048)}{375 + 450}$$

$$\hat{p} = 0.046$$

$$\hat{q} = 1 - p = 1 - 0.046 = 0.954$$

$$Z_{cal} = \frac{P_1 - P_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{0.045 - 0.048}{\sqrt{(0.046)(0.954)\left[\frac{1}{375} + \frac{1}{450}\right]}}$$

$$= -0.204$$

$$|Z_{cal}| = 0.204$$

Step-5 → $Z_{cal} < Z_{tab}$

We accept null hypothesis

NOTE: If only population proportions are given and sample proportions are not known, then the test static is

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

⑤ In 2 large populations, there are 30% & 25% of fair hair people. Is this difference likely to be hidden in samples of 1200 & 900 respectively from the 2 populations.

<u>sol:</u> Given $n_1 = 1200$ $n_2 = 900$

$$P_1 = 30\% = \frac{30}{100} = 0.3$$

$$P_2 = 25\% = \frac{25}{100} = 0.25$$

$$Q_1 = 1 - P_1 = 1 - 0.3 = 0.7$$
$$Q_2 = 1 - P_2 = 1 - 0.25 = 0.75$$

<u>Step-1</u> → $H_0 : P_1 = P_2$

<u>Step-2</u> → $H_1 : P_1 \neq P_2$ [Two-tailed Test]

<u>Step-3</u> → $\alpha = 0.05$

$Z_{tab} = 1.96$

<u>Step-4</u> →

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}}$$

$$= \frac{0.3 - 0.25}{\sqrt{\frac{0.3(0.7)}{1200} + \frac{0.25(0.75)}{900}}}$$

$Z_{cal} = 2.55$

<u>Step-5</u> → $Z_{cal} > Z_{tab}$

We reject null hypothesis

27/11/21

⑥ A random sample of 300 shops at a super market include 204 who regularly use coupons. In another sample of 500 shops at a super market includes 75 who regularly uses coupons. Construct 95% confidence interval for the probability that any one shop at the super market selected at random will regularly use coupons.

Sol: $n_1 = 300$ $x_1 = 204$

$n_2 = 500$ $x_2 = 75$

$P_1 = \frac{x_1}{n_1} = \frac{204}{300} = 0.68$ $Q_1 = 1 - P_1 = 1 - 0.68 = 0.32$

$P_2 = \frac{x_2}{n_2} = \frac{75}{500} = 0.15$ $Q_2 = 1 - P_2 = 1 - 0.15 = 0.85$

~~Step-1 ⇒ $H_0 : P_1 = P_2$~~

95% confidence interval ⇒ 5% LOS
= 0.05

$Z_{\alpha/2} = Z_\alpha = 1.96$

Confidence interval is $\left(P_1 - Z_{\alpha/2}\sqrt{\frac{P_1 q_1}{n_1}}, P_1 + Z_{\alpha/2}\sqrt{\frac{P_1 q_1}{n_1}}\right)$

$= \left(0.68 - 1.96\sqrt{\frac{0.68 \times 0.32}{300}}, 0.68 + 1.96\sqrt{\frac{0.68 \times 0.32}{300}}\right)$

$= (0.627, 0.732)$

⑦ The items produced by a factory out of 500, 15 were defective, in another sample out of 400, 20 were defective. Test the significance difference b/w 2 proportions at 5% LOS. Construct 95% confidence interval for the true proportion.

Sol: $n_1 = 500 \quad x_1 = 15$

$n_2 = 400 \quad x_2 = 20$

$$P_1 = \frac{x_1}{n_1} = \frac{15}{500} = 0.03$$

$$P_2 = \frac{x_2}{n_2} = \frac{20}{400} = 0.05$$

Step-1 → $H_0: P_1 = P_2$

Step-2 → $H_1: P_1 \neq P_2$ [Two-tailed Test]

Step-3 → $\alpha = 0.05$

$Z_{tab} = 1.96$

Step-4 → $$Z_{cal} = \frac{P_1 - P_2}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}}$$

$$\hat{P} = \frac{n_1P_1 + n_2P_2}{n_1 + n_2} = \frac{500(0.03) + 400(0.05)}{500 + 400}$$

$$= 0.038$$

$$\hat{q} = 1 - \hat{P} = 1 - 0.038$$

$$= 0.962$$

$$Z_{cal} = \frac{0.03 - 0.05}{\sqrt{(0.038)(0.962)\left[\frac{1}{500} + \frac{1}{400}\right]}}$$

$$= -1.55$$

$$|Z_{cal}| = |-1.55|$$

$$= 1.55$$

Step-5 → $Z_{cal} < Z_{tab}$

We accept null hypothesis

Confidence Interval

$P_1 = 0.03 \qquad Q_1 = 1 - P_1 = 1 - 0.03 = 0.97$

$P_2 = 0.05 \qquad Q_2 = 1 - P_2 = 1 - 0.05 = 0.95$

95% confidence interval = 5% LOS

$= 0.05$

$Z_\alpha = 1.96$

Confidence interval (single) $\left(P_1 - Z_\alpha\sqrt{\frac{P_1Q_1}{n_1}}, P_1 + Z_\alpha\sqrt{\frac{P_1Q_1}{n_1}}\right)$

$$= \left(0.03 - 1.96\sqrt{\frac{(0.03)(0.97)}{500}}, 0.03 + 1.96\sqrt{\frac{(0.03)(0.97)}{500}}\right)$$

$$= (0.01, 0.04)$$

Confidence interval

$$\left((P_1 - P_2) - Z_{\alpha/2}\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}, (P_1 - P_2) + Z_{\alpha/2}\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right)$$

$$= \left((0.03-0.05) - 1.96\sqrt{(0.038)(0.962)\left[\frac{1}{500}+\frac{1}{400}\right]}, (0.03-0.05) + 1.96\sqrt{(0.038)\times(0.962)\left[\frac{1}{500}+\frac{1}{400}\right]}\right)$$

$$= (-0.02, -0.01)$$

$$= (0.02, 0.01)$$

H.W

① Two types of new cars produced in USA are tested for petrol mileage, one sample is consisting of 42 cars averaged 15kmpl while the other sample consisting of 80 cars averaged 11.5 kmpl with population variances $\sigma_1^2 = 2$ and $\sigma_2^2 = 1.5$ respectively. Test whether there is any significant difference in petrol consumption of these 2 types of cars at 1% LOS (M-2)

② A manufacturer claims that only 4% of his products are defective. A random sample of 500 were taken among which 100 are defective. Test the hypothesis at 5% LOS (M-3)

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 at 5% LOS (M-1)

④ Among the items produced by a factory out of 800, 65 were defective. In another sample out of 300, 40 were defective. Test the significant difference at 1% LOS.

① $n_1 = 42$ $\quad n_2 = 80$

$\bar{x}_1 = 15$ $\quad \bar{x}_2 = 11.5$

$\sigma_1^2 = 2$ $\quad \sigma_2^2 = 1.5$

Step-1 → Set up null hypothesis $H_0 : \mu_1 = \mu_2$

Step-2 → Set up alternative hypothesis

$H_1 : \mu_1 \neq \mu_2$ [Two-tailed Test]

Step-3 → $\alpha = 0.01$

$Z_{tabulated} = 2.58$

Step-4 → $$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{15 - 11.5}{\sqrt{\frac{2}{42} + \frac{1.5}{80}}}$$

$Z_{calculated} = 13.58$

Step-5 → $Z_{calculated} > Z_{tabulated}$

∴ We reject null hypothesis

② $n = 500, x = 100$ $\quad$ ~~$P = \frac{1}{2}, Q = \frac{1}{2}$~~

$P = \frac{x}{n} = \frac{100}{500} = 0.2$ $\quad$ $P = 4\% = 0.04$; $Q = 1 - P = 1 - 0.04$

$Q = 0.96$

Step-1 → Set up null hypothesis $H_0 : P = P_0$

Step-2 → Set up alternative hypothesis

$H_1 : P \neq P_0$ [Two tailed Test]

Step-3 → $\alpha = 0.05$

$$Z_{tabulated} = 1.96$$

Step-4 → $$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.2-0.04}{\sqrt{\frac{0.04 \times 0.96}{500}}}$$

$$Z_{cal} = 18.25$$

Step-5 → $$Z_{calculated} > Z_{tabulated}$$

∴ We reject null hypothesis

(3) $n = 400$, $\sigma = 10$, $\bar{x} = 40$, $\mu = 38$

Step-1 → Set up null hypothesis : $\mu = \mu_0$

Step-2 → Set up alternative hypothesis H_1 : $\mu \neq \mu_0$ [Two-tailed Test]

Step-3 → $\alpha = 0.05$

$$Z_{tabulated} = 1.96$$

Step-4 → $$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

$$= \frac{40-38}{10/\sqrt{400}}$$

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

$$Z_{calculated} = 4$$

Step-5 → $Z_{calculated} > Z_{tabulated}$

∴ We reject null hypothesis

(4) $n_1 = 800 \qquad x_1 = 65$

$n_2 = 300 \qquad x_2 = 40$

$$P_1 = \frac{x_1}{n_1} = \frac{65}{800} = 0.08$$

$$P_2 = \frac{x_2}{n_2} = \frac{40}{300} = 0.13$$

Step-1 → Set up null hypothesis $H_0 : P_1 = P_2$

Step-2 → Set up alternative hypothesis $H_1 : P_1 \neq P_2$

[Two-tailed Test]

Step-3 → $\alpha = 0.01$

$Z_{tabulated} = Z_\alpha = 2.58$

Step-4 → $$Z_{cal} = \frac{P_1 - P_2}{\sqrt{\hat{p}\hat{q}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

$$\hat{P} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{800(0.08) + 300(0.13)}{800 + 300}$$

$$= 0.093$$

$$\hat{q} = 1 - \hat{P} = 1 - 0.093 = 0.907$$

$$Z_{cal} = \frac{0.08 - 0.13}{\sqrt{(0.093)(0.907)\left[\frac{1}{800} + \frac{1}{300}\right]}}$$

$$= -2.54$$

$$|Z_{cal}| = 2.54$$

Step-5 →

$$Z_{tabulated} > Z_{calculated}$$

∴ We accept null hypothesis

(OR)

$\hat{P} = 0.09$ $\hat{q} = 0.91$

$$Z = \frac{0.08 - 0.13}{\sqrt{(0.09)(0.91)\left[\frac{1}{800} + \frac{1}{300}\right]}}$$

$Z_{cal} = -2.58$

$|Z_{cal}| = 2.581$

$Z_{cal} > Z_{tab}$

∴ We reject null hypothesis

29/11/21

UNIT-IV

Test of Significance for Small Samples

<u>Small Sample</u>: If the size of the sample i.e $n \leq 30$, then the samples are said to be small samples.

* Under small sample test, we have 3 tests

a) t-Test (or) Student t-test

b) F-Test

c) Chi-Square Test [χ^2-Test]

1. **Student t-test**: Testing small samples using t-distribution is called t-test.

Assumptions:

a) The size of the sample should be $n \leq 30$.

b) The parent population from which sample is drawn is normal.

c) The population S.D is unknown.

d) The sample observations are independent.

Uses:

a) To test for a specified mean.

b) To test for equality of two means of two independent samples drawn from two normal populations, S.D of population is unknown.

c) To test the significance of difference between the means of paired data.

Properties:

a) The shape of t-distribution is bell-shaped which is similar to a normal distribution and is symmetrical about the mean.

b) It is symmetrical about the line $t=0$.

c) The form of the probability curves varies with degrees of freedom i.e with sample size.

d) It is unimodel i.e Mean = Median = Mode

e) The mean of Standard Normal Distribution and t-distribution is zero but the variance of t-distribution depends upon the parameter 'v' which is called Degrees of Freedom.

f) The variance of t-distribution exceeds 1, but approaches 1 as $t \to \infty$

Procedure:

Step-1 → Set up null hypothesis i.e $H_0: \mu = \mu_0$

tep-2 → Set up alternative hypothesis i.e

$H_1: \mu \neq \mu_0$ [Two-tailed Test]

$H_1: \mu > \mu_0$ [Right-tailed Test]

$H_1: \mu < \mu_0$ [Left-tailed Test]

Step-3 → Write LOS from the given problem, if not take $\alpha = 0.05$

Step-4 → Calculate the test statistic

The test statistic for t-test (Single Mean) is

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} \quad [\text{SD is known}]$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad [\text{SD is unknown}]$$

Step-5 → Find the degrees of freedom i.e $V = n-1$ corresponding to LOS α value from t-distribution table i.e $t_{tabulated}$ value.

Compare $t_{calculated}$ & $t_{tabulated}$ Values

a) If $t_{calculated} < t_{tabulated}$, then we accept null hypothesis

b) If $t_{calculated} > t_{tabulated}$, then we reject null hypothes-is

NOTE: The t-distribution table values are given for one-tailed test. So, in case of two-tailed test Consider LOS as $\frac{\alpha}{2}$ i.e ~~one~~ two-tailed

α i.e one-tailed

Degrees of Freedom: The no. of independent values or quantities which can be assigned to a statistical distribution i.e $\boxed{V = n - K}$

① A sample of 26 bulbs gives a mean life of 990 hrs with a S.D of 20 hrs. The manufacturer claims that the mean life of the bulbs is 1000 hrs. Is the sample not upto the standard.

Sol: Given $n = 26$

$\mu = 1000$ hrs

$\bar{x} = 990$ hrs (Sample)

$s = 20$ hrs

Step-1 → Set up null hypothesis i.e

$$H_0 : \mu = \mu_0 = 1000$$

Step-2 → Set up alternative hypothesis i.e

$$H_1 : \mu \neq 1000 \quad [\text{Two-tailed Test}]$$

Step-3 → $\alpha = 0.05$

$$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$

Step-4 → Calculate the test statistic i.e

$$t_{calculated} = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

$$= \frac{990 - 1000}{20/\sqrt{26-1}}$$

$$= \frac{-10}{\frac{204}{81}}$$

$$= -2.5$$

$$|t_{calculated}| = 2.5$$

Step-5→ Find degrees of freedom from t-distribution table corresponding to $\frac{\alpha}{2}$ value.

$V = n-1$

$= 26-1$

$V = 25$, $\frac{\alpha}{2} = 0.025$

$t_{tabulated} = 2.060$

Compare $t_{calculated}$ & $t_{tabulated}$ values

$t_{calculated} > t_{tabulated}$

∴ We reject null hypothesis

② The average breaking strength of the steel rods is specified to be 18.5 pounds. To test this sample of 14 rods were tested. The mean & S.D obtained as 17.85 and 1.955. Is the result of the experiment is significant.

Sol:- $n = 14$

$\bar{x} = 17.85$

$\mu = 18.5$

$s = \sigma = 1.955$

Step-1 → $H_0 : \mu = \mu_0$

Step-2 → $H_1 : \mu \neq \mu_0$ [Two-tailed Test]

Step-3 → $\alpha = 0.05$

$$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$

Step-4 →

$$t_{calculated} = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

$$= \frac{17.85 - 18.5}{1.955/\sqrt{14-1}}$$

$$= -1.19$$

$$|t_{calculated}| = 1.19$$

Step-5 →

$V = n-1$ $\qquad \frac{\alpha}{2} = 0.025$

$= 14-1$

$= 13$

$$t_{tabulated} = 2.160$$

$t_{tabulated} > t_{calculated}$

$\therefore$ We accept null hypothesis

③ A random sample of 6 steel beams has a mean comprehensive strength of 58,392 P.S.I with a S.D of 648 P.S.I. Use this information & LOS $\alpha = 0.05$ to test whether the true average comprehensive strength of the steel which this sample came is 58,000 P.S.I.

Sol: $n = 6$

$\mu = 58000$

$\bar{x} = 58,392$

$s = 648$

Step-1 → $H_0 : \mu = \mu_0$

Step-2 → $H_1 : \mu \neq \mu_0$ [Two-tailed Test]

Step-3 → $\alpha = 0.05$

$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

Step-4 →

$$t_{calculated} = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

$$= \frac{58,392 - 58000}{648/\sqrt{6-1}}$$

$$t_{calculated} = 1.35$$

Step-5 →

DOF: $V = n-1$

$= 6-1$

$= 5$

$\frac{\alpha}{2} = 0.025$

$t_{tabulated} = 2.571$

$t_{tabulated} > t_{calculated}$

$\therefore$ We accept null hypothesis

$\boxed{30/11/21}$

Problems related to t-test when standard deviation of the sample is not given directly:

① A random sample of 10 boys have the following IQ's: 70, 120, 110, 101, 88, 83, 95, 98, 107 and 100.

a) Do this data support the assumption of a population Mean IQ of 100

b) Find a reasonable range in which most of the mean IQ values of samples of 10 boys lies.

Sol: Here, the mean & S.D of the sample is not given directly.

We have to determine mean & S.D as follows

$$\text{Mean } \bar{x} = \frac{70+120+110+101+88+83+95+98+107+100}{10}$$

$$= 97.2$$

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

x	$(x-\bar{x})$	$(x-\bar{x})^2$
70	70-97.2	739.84
120	120-97.2	519.84
110	110-97.2	163.84
101	101-97.2	14.44
88	88-97.2	84.64
83	83-97.2	201.64
95	95-97.2	4.84
98	98-97.2	0.64
107	107-97.2	96.04
100	100-97.2	7.84
		1833.6

$$s^2 = \frac{1}{10-1}(1833.6)$$

$$= \frac{1833.6}{9}$$

$$= 203.73$$

$$s = \sqrt{203.73} = 14.27$$

$n = 10,\ \mu = 100,\ \bar{x} = 97.2,\ s = 14.27$

Step-1 → Set up null hypothesis

$$H_0: \mu = 100$$

Step-2 → Set up alternative hypothesis

$$H_1: \mu \neq 100 \text{ [Two-tailed Test]}$$

Step-3 → $\alpha = 0.05$

$$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$

Step-4 → Calculate the test statistic

$$t_{cal} = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

$$= \frac{97.2-100}{\frac{14.27}{\sqrt{10}}}$$

$$t_{cal} = -0.62$$

$$|t_{cal}| = 0.62$$

Step-5 →

DOF $v = n-1$ $\quad \frac{\alpha}{2} = 0.025$

$= 10-1$

$= 9$

$$t_{tab} = 2.262$$

$$t_{tab} > t_{cal}$$

∴ We accept null hypothesis

b) The confidence interval for single mean is

$$\left(\bar{x} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right), \bar{x} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)\right)$$

$$= \left(97.2 - 2.262\left[\frac{14.27}{\sqrt{10}}\right], 97.2 + 2.262\left[\frac{14.27}{\sqrt{10}}\right]\right)$$

$$= (86.99, 107.40)$$

$$= (87, 107.40)$$

(2) The lifetime of electric bulbs for a random sample of 10 from a factory given the following data

Item	1	2	3	4	5	6	7	8	9	10
Life in 1000 hrs	1.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

Can we accept the hypothesis that the avg. lifetime of electric bulbs is 4000 hrs. Use 0.05 LOS

sol: Mean $\bar{x} = \dfrac{1.2+4.6+3.9+4.1+5.2+3.8+3.9+4.3+4.4+5.6}{10}$

$= \dfrac{41}{10}$

$\bar{x} = 4.1 \times 1000 \Rightarrow \bar{x} = 4100$

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

x	$x-\bar{x}$	$(x-\bar{x})^2$
1.2	1.2−4.1	8.41
4.6	4.6−4.1	0.25
3.9	3.9−4.1	0.04
4.1	4.1−4.1	0
5.2	5.2−4.1	1.21
3.8	3.8−4.1	0.09
3.9	3.9−4.1	0.04
4.3	4.3−4.1	0.04
4.4	4.4−4.1	0.09
5.6	5.6−4.1	2.25
		12.42

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

$= \frac{1}{10-1}(12.42)$

$= \frac{12.42}{9} \quad \Rightarrow s^2 = 1.38 \times 1000 \Rightarrow s^2 = 1380$

$S = \sqrt{1380} = 37.14$

$\mu = 4000, n = 10, \bar{x} = 4100, S = 37.14$

Step-1 → Set up null hypothesis $H_0: \mu = 4000$

Step-2 → Set up alternative hypothesis

$H_1: \mu \neq 4000$ [Two-tailed Test]

Step-3 → $\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

Step-4 → Calculate the test statistic

$$t_{cal} = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

$$= \frac{4100 - 4000}{\frac{37.14}{\sqrt{10}}}$$

$$t_{cal} = 8.51$$

Step-5 →

DOF $V = n - 1$ $\qquad \frac{\alpha}{2} = 0.025$

$= 10 - 1$

$= 9$

$t_{tabulated} = 2.262$

$t_{tabulated} < t_{calculated}$

$\therefore$ We reject null hypothesis

③ A random sample of 10 bags of pesticides are taken whose weights are 50, 49, 52, 44, 45, 48, 46, 45, 49, 45 Test whether the average packing can be taken to be 50kgs.

Sol:

$$\text{Mean } \bar{x} = \frac{50+49+52+44+45+48+46+45+49+45}{10}$$

$$= \frac{473}{10}$$

$$= 47.3$$

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

x	$x-\bar{x}$	$(x-\bar{x})^2$
50	50-47.3	7.29
49	49-47.3	2.89
52	52-47.3	22.09
44	44-47.3	10.89
45	45-47.3	5.29
48	48-47.3	0.49
46	46-47.3	1.69
45	45-47.3	5.29
49	49-47.3	2.89
45	45-47.3	5.29
		64.1

$$s^2 = \frac{1}{10-1}(64.1)$$

$$= \frac{64.1}{9}$$

$$s^2 = 7.12$$

$$s = \sqrt{7.12}$$

$$\Rightarrow s = 2.66,\ \bar{x} = 47.3,\ \mu = 50,\ n = 10$$

Step-1 → Set up null hypothesis $H_0: \mu = 50$

Step-2 → Set up alternative hypothesis $H_1: \mu \neq 50$ [Two tailed Test]

Step-3 → $\alpha = 0.05$

$$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$

Step-4 →

$$t_{cal} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{47.3 - 50}{2.64/\sqrt{10}}$$

$$= -3.20$$

$$|t_{cal}| = 3.2$$

Step-5 →

DOF $v = n - 1$ $\qquad \frac{\alpha}{2} = 0.025$

$= 10 - 1$

$= 9$

$t_{tab} = 2.262$

$t_{cal} > t_{tab}$

∴ We reject null hypothesis

(4) 8 students were given a test in statistics and after
ne month of coaching they were given another test.
he following table gives the increase in marks in
d test over the 1^{st}.

Student no	1	2	3	4	5	6	7	8
Increase of Marks	4	-2	6	-8	12	5	-7	2

Assume that the students are not benefitted by coaching, it implies that the mean of the difference of 2 tests is 0.

Sol:- Mean $\bar{x} = \frac{4-\cancel{2}+6-8+12+5-7+\cancel{2}}{8}$

$= \frac{12}{8}$

$= 1.5$

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

x	$x-\bar{x}$	$(x-\bar{x})^2$
4	4 − 1.5	6.25
−2	−2 − 1.5	12.25
6	6 − 1.5	20.25
−8	−8 − 1.5	90.25
12	12 − 1.5	110.25
5	5 − 1.5	12.25
−7	−7 − 1.5	72.25
2	2 − 1.5	0.25
		324

$s^2 = \frac{1}{8-1}(324)$

$s^2 = 46.28$

$s = \sqrt{46.28} \Rightarrow s = 6.80$

$\bar{x} = 1.5, \mu = 0, n = 8$

Step-1 → Set up null hypothesis $H_0: \mu = 0$

Step-2 → Set up alternative hypothesis $H_1: \mu > 0$ [Right tailed]

Step-3 → $\alpha = 0.05$

Step-4 → Calculate the test statistic

$$t_{cal} = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

$$= \frac{1.5 - 0}{\frac{6.8}{\sqrt{8}}}$$

$$t_{cal} = 0.62$$

Step-5 →

DOF $V = n - 1$ $\qquad \alpha = 0.05$

$= 8 - 1$

$= 7$

$t_{tab} = 1.895$

$t_{tab} > t_{cal}$

$\therefore$ We accept null hypothesis

01/12/21

Student t-test for difference of means: Let $\bar{x}$ & $\bar{y}$ be the means of two independent samples of sizes n_1 & n_2 drawn from two normal populations having means μ_1 & μ_2. To test the significance for the difference of 2 means, we follow the following steps.

Step-1 → Set up null hypothesis $H_0: \mu_1 = \mu_2$

Step-2 → Set up alternative hypothesis

$H_1: \mu_1 \neq \mu_2$ [Two-tailed]

$H_1: \mu_1 > \mu_2$ [Right-tailed]

$H_1: \mu_1 < \mu_2$ [Left-tailed]

Step-3 → Write LOS α for the given problem, if not take $\alpha = 0.05$

Step-4 → Calculate the test statistic

The test statistic for difference of two means is

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

where

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

(If S.D is known)

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2\right]$$

(If S.D is unknown)

Step-5 → Find the degrees of freedom i.e $V = n_1 + n_2 - 2$ and find tabulated value from t-distribution table corresponding to degrees of freedom & LOS α

Compare t_{cal} & t_{tab} values

a) If $t_{cal} < t_{tab}$, we accept null hypothesis

b) If $t_{cal} > t_{tab}$, we reject null hypothesis

① Samples of 2 types of electric bulbs are tested & following data is obtained

Type - I	Type - II
$n_1 = 8$	$n_2 = 7$
$\bar{x} = 1234$ hrs	$\bar{y} = 1036$ hrs
$s_1 = 36$ hrs	$s_2 = 40$ hrs

Is the difference in the means sufficient to warranty that type-I is superior to type-II regarding length of life.

Sol:

Step-1 → Set up null hypothesis H_0: There is no difference between two types of bulbs

$$H_0 : \mu_1 = \mu_2$$

Step-2 → Set up alternative hypothesis i.e $H_1 : \mu_1 > \mu_2$ [Right-tailed Test]

Step-3 → $\alpha = 0.05$

Step-4 → Calculate the test statistic

$$t_{cal} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2} = \frac{8(36)^2 + 7(40)^2}{8+7-2}$$

$$= \frac{21568}{13}$$

$$s^2 = 1659.07$$

$$t_{cal} = \frac{1234-1036}{\sqrt{(1659.07)\left[\frac{1}{8}+\frac{1}{7}\right]}}$$

$$= \frac{198}{\sqrt{(1659.07)\left(\frac{1}{8}+\frac{1}{7}\right)}}$$

$$t_{cal} = 9.39$$

Step-5 →

$\alpha = 0.05$

$$DOF = n_1+n_2-2$$
$$= 8+7-2$$
$$= 13$$

$$t_{tab} = 1.771$$

Compare t_{cal} & t_{tab} Values

$t_{cal} > t_{tab}$

∴ We reject null hypothesis

Type-I is superior to Type-II

② 2 independent samples of 8 & 7 items respectively had the following values

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

Is the difference between the means of the sample significant or not.

Sol: $n_1 = 8$ $n_2 = 7$

$$\bar{x} = \frac{11+11+13+11+15+9+12+14}{8}$$

$$\bar{x} = 12$$

$$\bar{y} = \frac{9+11+10+13+9+8+10}{7} = \frac{70}{7}$$

$$\bar{y} = 10$$

$$S^2 = \frac{1}{n_1+n_2-2}\left[\sum(x_i-\bar{x})^2 + \sum(y_i-\bar{y})^2\right]$$

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
11	11 – 12	1
11	11 – 12	1
13	13 – 12	1
11	11 – 12	1
15	15 – 12	9
9	9 – 12	9
12	12 – 12	0
14	14 – 12	4
		26

y_i	$y_i-\bar{y}$	$(y_i-\bar{y})^2$
9	9 – 10	1
11	11 – 10	1
10	10 – 10	0
13	13 – 10	9
9	9 – 10	1
8	8 – 10	4
10	10 – 10	0
		16

$$s^2 = \frac{1}{8+7-2}\left[26+16\right]$$

$$s^2 = 3.23$$

Step-1 → $H_0: \mu_1 = \mu_2$

Step-2 → $H_1: \mu_1 \neq \mu_2$ [Two-tailed Test]

Step-3 → $\alpha = 0.05$

$$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$

Step-4 →

$$t_{cal} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{12 - 10}{\sqrt{(3.23)\left(\frac{1}{8} + \frac{1}{7}\right)}}$$

$$t_{cal} = 2.15$$

Step-5 →

DOF $\quad v = n_1 + n_2 - 2 \qquad \frac{\alpha}{2} = 0.025$

$= 8 + 7 - 2$

$= 13$

$$t_{tab} = 2.160$$

$$t_{cal} < t_{tab}$$

∴ We accept null hypothesis

③ The intelligence quotient of 16 students from one area of a city has a mean 107 with a S.D 10, while the IQ's of 14 students from another area of the city has a mean 112 with a S.D of 8. Is there any significant difference between the IQ's of 2 groups at 0.05 LOS.

Sol: Given $n_1 = 16$ $\quad n_2 = 14$

$\bar{x} = 107 \quad \bar{y} = 112$

$s_1 = 10 \quad s_2 = 8$

Step-1 → $H_0: \mu_1 = \mu_2$

Step-2 → $H_1: \mu_1 \neq \mu_2$ [Two-tailed Test]

Step-3 → $\alpha = 0.05$

$$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$

Step-4 →

$$t_{cal} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{16(10)^2 + 14(8)^2}{16 + 14 - 2} = \frac{2496}{28} = 89.14$$

$$t_{cal} = \frac{107 - 112}{\sqrt{(89.14)\left[\frac{1}{16} + \frac{1}{14}\right]}}$$

$$= -1.44$$

$$|t_{cal}| = 1.44$$

Step-5 →

DOF $V = n_1 + n_2 - 2$ $\qquad \frac{\alpha}{2} = 0.025$

$= 16 + 14 - 2$

$= 28$

$t_{tab} = 2.048$

$t_{cal} < t_{tab}$

$\therefore$ We accept null hypothesis

(4) To examine the hypothesis that the husbands are more intelligent than the wives, an investigator took a sample of 10 couples & tested the IQ's. The results are as follows

Husband	117	105	97	105	123	109	86	78	103	107
Wife	106	98	87	104	116	95	90	69	108	85

Test the hypothesis at 5% LOS

Sol:- $\quad n_1 = 10 \qquad n_2 = 10$

$$\bar{x} = \frac{117+105+97+105+123+109+86+78+103+107}{10}$$

$$= \frac{1030}{10}$$

$$\bar{x} = 103$$

$$\bar{y} = \frac{106+98+87+104+116+95+90+69+108+85}{10}$$

$$= \frac{958}{10}$$

$$= 95.8$$

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
117	117−103	196
105	105−103	4
97	97−103	36
105	105−103	4
123	123−103	400
109	109−103	36
86	86−103	289
78	78−103	625
103	103−103	0
107	107−103	16
		1606

y	$y_i-\bar{y}$	$(y_i-\bar{y})^2$
106	106−95.8	104.04
98	98−95.8	4.84
87	87−95.8	77.44
104	104−95.8	67.24
116	116−95.8	408.04
95	95−95.8	0.64
90	90−95.8	33.64
69	69−95.8	718.24
108	108−95.8	148.84
85	85−95.8	116.64
		1679.6

$$S^2 = \frac{1}{n_1+n_2-2}\left[\sum(x_i-\bar{x})^2 + \sum(y_i-\bar{y})^2\right]$$

$$= \frac{1}{10+10-2}\left[1606 + 1679.6\right]$$

$$= \frac{3285.6}{18}$$

$$= 182.53$$

Step-1 → $H_0: \mu_1 = \mu_2$

Step-2 → $H_1: \mu_1 > \mu_2$ [Right-tailed Test]

Step-3 → $\alpha = 0.05$

Step-4 →

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$t = \frac{103 - 95.8}{\sqrt{182.53\left[\frac{1}{10}+\frac{1}{10}\right]}}$$

$$= \frac{7.2}{\sqrt{36.506}}$$

$$t_{calculated} = 1.191$$

Step-5 →

DOF $v = n_1 + n_2 - 2$ $\alpha = 0.05$

$= 10 + 10 - 2$

$= 18$

$$t_{tabulated} = 1.734$$

$$t_{calculated} < t_{tabulated}$$

∴ We accept null hypothesis

04/12/21

F_test (or) F-distribution: When testing the significance of difference of means of 2 samples we assume that two samples came from the same population or from populations with equal variances. If the variances of the population are not equal, a significant difference in the mean may occur. Hence, before we apply t-test for the significance of difference of 2 means, we have to test for the equality of population

Variances using F-test.

NOTE:

① t-test is used to test the significance of difference of two means.

② F-test is used to test the significance difference between two variances.

Procedure:

Step-1 → Set up null hypothesis

$$H_0 : \sigma_1^2 = \sigma_2^2$$

Step-2 → Set up alternative hypothesis

$$H_1 : \sigma_1^2 \neq \sigma_2^2 \text{ [Two-tailed]}$$

$$H_1 : \sigma_1^2 > \sigma_2^2 \text{ [Right-tailed]}$$

$$H_1 : \sigma_1^2 < \sigma_2^2 \text{ [Left-tailed]}$$

Step-3 → Write LOS α from given problem if not take $\alpha = 0.05$

Step-4 → Calculate the test statistic

The test statistic for F-test is

$$F = \frac{\text{Greater Variance}}{\text{Smaller Variance}}$$

If $S_2^2 > S_1^2$ interchange v_1 & v_2

i.e $F = \dfrac{S_1^2}{S_2^2}$ $\left[\text{if } S_1^2 > S_2^2\right]$ (v_1, v_2) (or) n_1 & n_2

$F = \dfrac{S_2^2}{S_1^2}$ $\left[\text{if } S_2^2 > S_1^2\right]$ (v_2, v_1)

where $$\left.\begin{aligned} S_1^2 &= \frac{1}{n_1-1}\sum(x_i-\bar{x})^2 \\ S_2^2 &= \frac{1}{n_2-1}\sum(y_i-\bar{y})^2 \end{aligned}\right\}$$ If sum of squares of deviation is known

$$\left.\begin{aligned} S_1^2 &= \frac{n_1 s_1^2}{n_1-1} \\ S_2^2 &= \frac{n_2 s_2^2}{n_2-1} \end{aligned}\right\}$$ If only S.D of sample is known

Step-5 → Find degrees of freedom i.e

$V_1 = n_1 - 1 \qquad V_2 = n_2 - 1$

Find the tabulated value from f-distribution table corresponding to degrees of freedom v_1 & v_2 and LOS α

- If $F_{calculated} < F_{tabulated}$, then we accept null hypothesis
- If $F_{calculated} > F_{tabulated}$ then we reject null hypothes[is]

① In one sample of 8 observations from a norma[l] Population, the sum of squares of deviation of the sample values from sample mean is 84.4 & in another sample of 10 observations it was 102.6 At 5% LOS, test whether the populations have same Variance

Sol: Given $n_1 = 8$ $n_2 = 10$

$$\Sigma(x_i - \bar{x})^2 = 84.4 \qquad \Sigma(y_i - \bar{y})^2 = 102.6$$

Step-1 → Set up null hypothesis

$$H_0: \sigma_1^2 = \sigma_2^2$$

Step-2 → Set up alternative hypothesis

$$H_1: \sigma_1^2 \neq \sigma_2^2 \text{ [Two-tailed Test]}$$

Step-3 → $\alpha = 0.05$

Step-4 → Calculate the test statistic i.e

$$S_1^2 = \frac{1}{n_1 - 1} \Sigma(x_i - \bar{x})^2$$

$$= \frac{1}{8-1}(84.4)$$

$$= 12.05$$

$$S_2^2 = \frac{1}{n_2 - 1} \Sigma(y_i - \bar{y})^2$$

$$= \frac{1}{10-1}(102.6)$$

$$= 11.4$$

Here, $S_1^2 > S_2^2$

$$\therefore F_{calculated} = \frac{S_1^2}{S_2^2}$$

$$= \frac{12.05}{11.4}$$

$$= 1.057$$

Step-5 →

$V_1 = n_1 - 1$ $\quad$ $V_2 = n_2 - 1$ $\quad$ $\alpha = 0.05$

$= 8 - 1$ $\quad$ $= 10 - 1$ $\quad$ $V_1, V_2 = 7, 9$

$= 7$ $\quad$ $= 9$

$F_{tabulated} = 3.29$ (From f-distribution table)

Compare $F_{calculated}$ & $F_{tabulated}$ Values

$F_{calculated} < F_{tabulated}$

∴ We accept null hypothesis

2) Pumpkins are grown under 2 experimental conditions. Two random samples of 11 & 9 pumpkins show the sample S.D as 0.8 & 0.5 respectively. Test the hypothesis that the variances are equal.

Sol: Given $n_1 = 11$ $\quad$ $n_2 = 9$

$s_1 = 0.8$ $\quad$ $s_2 = 0.5$

Step-1 → $H_0: \sigma_1^2 = \sigma_2^2$

Step-2 → $H_1: \sigma_1^2 \neq \sigma_2^2$ [Two-tailed Test]

Step-3 → $\alpha = 0.05$

Step-4 →

$$F_{calculated} = \frac{\text{Greater Variance}}{\text{Smaller Variance}}$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1}$$

$$= \frac{11(0.8)^2}{11-1}$$

$$= 0.704$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1}$$

$$= \frac{9(0.5)^2}{9-1}$$

$$= 0.281$$

$$S_1^2 > S_2^2$$

$$F_{calculated} = \frac{S_1^2}{S_2^2}$$

$$= \frac{0.704}{0.281}$$

$$= 2.505$$

Step-5 →

$v_1 = n_1 - 1$
$= 11 - 1$
$= 10$

$v_2 = n_2 - 1$
$= 9 - 1$
$= 8$

$\alpha = 0.05$

$$F_{tabulated} = 3.35$$

$$F_{calculated} < F_{tabulated}$$

∴ We accept null hypothesis

③ Time taken by the workers in performing a job Method-I and Method-II is given below

Method-I	20	16	26	27	23	22	–
Method-II	27	33	42	35	32	34	38

Test the hypothesis at 5% Los.

Sol:-

$$\bar{x} = \frac{20+16+26+27+23+22}{6} \qquad n_1 = 6$$

$$\bar{x} = 22.33$$

$$\bar{y} = \frac{27+33+42+35+32+34+38}{7} \qquad n_2 = 7$$

$$\bar{y} = 34.4$$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
20	20 -22.3	5.29
16	16-22.3	39.69
26	26 -22.3	13.69
27	27-22.3	22.09
23	23-22.3	0.49
22	22-22.3	0.09
		81.34

y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
27	27 -34.4	54.76
33	33-34.4	1.96
42	42-34.4	57.76
35	35-34.4	0.36
32	32-34.4	5.76
34	34 -34.4	0.16
38	38 -34.4	12.96
		133.72

Step-1 → $H_0 : \sigma_1^2 = \sigma_2^2$

Step-2 → $H_1 : \sigma_1^2 \neq \sigma_2^2$ [Two-tailed Test].

Step-3 → $\alpha = 0.05$

Step-4 →

$$S_1^2 = \frac{1}{n_1-1} \Sigma(x_i - \bar{x})^2 \qquad S_2^2 = \frac{1}{n_2-1} \Sigma(y_i - \bar{y})^2$$

$$= \frac{1}{6-1}(81.34) \qquad = \frac{1}{7-1}(133.72)$$

$$= 16.26 \qquad = 22.28$$

$S_2^2 > S_1^2$

$$F_{calculated} = \frac{S_2^2}{S_1^2}$$

$$= \frac{22.28}{16.26}$$

$$= 1.370$$

Step-5 →

$v_1 = n_1 - 1$
$= 6 - 1$
$= 5$

$v_2 = n_2 - 1$
$= 7 - 1$
$= 6$

$\alpha = 0.05$

$F_{tab} = 4.39$

$F_{calculated} < F_{tabulated}$

∴ We accept null hypothesis

06/12/21

(14) Two random samples gives the following results.

Sample	Size	Sample Mean	Sum of Squares of deviation
1	10	15	90
2	12	14	108

Test whether the samples came from the same normal population.

Q: If mean & sum of squares of deviations are given we have to apply

a) F-test b) T-test

Given $n_1 = 10$ $n_2 = 12$

$\bar{x} = 15$ $\bar{y} = 14$

$\sum(x_i - \bar{x})^2 = 90$ $\sum(y_i - \bar{y})^2 = 108$

a) F-test:

Step-1 → $H_0: \sigma_1^2 = \sigma_2^2$

Step-2 → $H_1: \sigma_1^2 \neq \sigma_2^2$ [Two-tailed Test]

Step-3 → $\alpha = 0.05$

Step-4 →

$$F_{calculated} = \frac{S_1^2}{S_2^2} \text{ or } \frac{S_2^2}{S_1^2}$$

$$S_1^2 = \frac{1}{n_1 - 1} \sum(x_i - \bar{x})^2$$

$$= \frac{1}{10-1}(90)$$

$$= \frac{90}{9}$$

$$= 10$$

$$S_2^2 = \frac{1}{n_2 - 1} \sum(y_i - \bar{y})^2$$

$$= \frac{1}{12-1}(108)$$

$$= \frac{108}{11}$$

$$= 9.818$$

$$S_1^2 > S_2^2$$

$$F_{calculated} = \frac{S_1^2}{S_2^2} = \frac{10}{9.818}$$

$$F_{cal} = 1.018$$

Step-5 →

$\alpha = 0.05$

$V_1 = n_1 - 1$ $\quad$ $V_2 = n_2 - 1$

$= 10 - 1$ $\quad$ $= 12 - 1$

$= 9$ $\quad$ $= 11$

$F_{tabulated} = 2.90$

$F_{calculated} < F_{tabulated}$

$\therefore$ We accept null hypothesis

b) T-test:

Step-1 → $H_0 : \mu_1 = \mu_2$

Step-2 → $H_1 : \mu_1 \neq \mu_2$ [Two-tailed Test]

Step-3 → $\alpha = 0.05$

Step-4 →

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2\right]$$

$$= \frac{1}{10 + 12 - 2}\left[90 + 108\right]$$

$$= \frac{198}{20}$$

$$= 9.9$$

$$F_{calculated} = \frac{\bar{x} - \bar{y}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{15 - 14}{\sqrt{9.9\left(\frac{1}{10} + \frac{1}{12}\right)}}$$

$$t_{calculated} = \frac{1}{1.3472}$$

$$= 0.742$$

Step-5 →

$v_0 = n_1 + n_2 - 2$ $\qquad\qquad \frac{\alpha}{2} = 0.025$

$= 10 + 12 - 2$

$= 20$

$t_{tabulated} = 2.086$

$t_{calculated} < t_{tabulated}$

$\therefore$ We accept null hypothesis

Since in both the tests we accept null hypothesis

$\therefore$ We can conclude that the given samples have been drawn from the same normal population.

⑤ The reporting levels in milligrams in 2 samples of tobacco are found to be as follows

Sample-A	24	27	26	21	25	-
Sample-B	27	30	28	31	22	36

Can it be said that the two samples have come from the same normal population.

Sol: $\qquad n_1 = 5 \qquad\qquad n_2 = 6$

$$\bar{x} = \frac{24+27+26+21+25}{5} \qquad \bar{y} = \frac{27+30+28+31+22+36}{6}$$

$$= \frac{123}{5} \qquad\qquad = \frac{174}{6}$$

$$= 24.6 \qquad\qquad = 29$$

x	$x-\bar{x}$	$(x-\bar{x})^2$
24	24-24.6	0.36
27	27-24.6	5.76
26	26-24.6	1.96
21	21-24.6	12.96
25	25-24.6	0.16
		21.2

y	$y-\bar{y}$	$(y-\bar{y})^2$
27	27-29	4
30	30-29	1
28	28-29	1
31	31-29	4
22	22-29	49
36	36-29	49
		108

a) F-Test:

Step-1 → $H_0: \sigma_1^2 = \sigma_2^2$

Step-2 → $H_1: \sigma_1^2 \neq \sigma_2^2$ [Two-tailed Test]

Step-3 → $\alpha = 0.05$

Step-4 →

$$S_1^2 = \frac{1}{n_1-1} \Sigma(x_i-\bar{x})^2$$

$$= \frac{1}{5-1} (21.2)$$

$$= \frac{21.2}{4}$$

$$= 5.3$$

$$S_2^2 = \frac{1}{n_2-1} \Sigma(y_i-\bar{y})^2$$

$$= \frac{1}{6-1} (108)$$

$$= \frac{108}{5}$$

$$= 21.6$$

$$S_2^2 > S_1^2$$

$$F_{calculated} = \frac{S_2^2}{S_1^2}$$

$$= \frac{21.6}{5.3}$$

$$= 4.075$$

Step-5 →

$\nu_2 = n_1 - 1$ $\quad\quad$ $\nu_1 = n_2 - 1$ $\quad\quad$ $\alpha = 0.05$

$= 5-1$ $\quad\quad$ $= 6-1$

$= 4$ $\quad\quad$ $= 5$

$F_{tabulated} = \cancel{5.19}\ 6.26$

$F_{calculated} < F_{tabulated}$

∴ We accept null hypothesis

(b) T-test:

Step-1 → $H_0 : \mu_1 = \mu_2$

Step-2 → $H_1 : \mu_1 \neq \mu_2$ [Two-tailed Test]

Step-3 → $\alpha = 0.05$ $\quad\quad$ $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

Step-4 →

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2\right]$$

$$= \frac{1}{5+6-2}\left[21.2 + 108\right]$$

$$= \frac{129.2}{9}$$

$$= 14.35$$

$$t_{calculated} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{24.6 - 29}{\sqrt{(14.35)\left[\frac{1}{5} + \frac{1}{6}\right]}} = -1.901$$

$$|t_{calculated}| = 1.901$$

Step-5 →

$\frac{\alpha}{2} = 0.025$

$V = n_1 + n_2 - 2$

$= 5 + 6 - 2$

$= 9$

$t_{tabulated} = 2.262$

$t_{calculated} < t_{tabulated}$

∴ We accept null hypothesis

Since in both tests we accept null hypothesis,

∴ The samples have been drawn from same normal Population.

Chi-Square Test (χ^2-test) for independence of attributes

* Literally, an attribute means a quality or characteristic

Eg: Drinking, Smoking, Blindness, Honesty, Beauty etc

* An attribute may be marked by its presence or absence in a number of a given population.

* Let us consider two attributes A & B. A is divided into two classes and B is divided into two classes. The various frequencies can be expressed in the following table

Attri-bute			
A	a	b	a+b
B	c	d	c+d
	a+c	b+d	a+b+c+d = N

* The above frequencies are known as observed frequencies.

For these observed frequencies corresponding expected frequencies are given in the following table.

$E(a) = \frac{(a+c)(a+b)}{N}$	$E(b) = \frac{(b+d)(a+b)}{N}$
$E(c) = \frac{(a+c)(c+d)}{N}$	$E(d) = \frac{(b+d)(c+d)}{N}$

Procedure:

Step-1 → Set up null hypothesis

Step-2 → Set up alternative hypothesis

Step-3 → Write LOS α from the given problem if not take $\alpha = 0.05$

Step-4 → Calculate the test statistic i.e

$$\chi^2 = \frac{(O_i - E_i)^2}{E_i}$$

where $O_i \rightarrow$ Observed Frequency

$E_i \rightarrow$ Expected Frequency

i.e $E_i = \dfrac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$

Step-5 → Find degrees of freedom

dof = (No. of rows - 1) (No. of columns - 1)

Corresponding to dof & LOS α value find the tabulated value from Chi-square distribution table.

Compare $\chi^2_{calculated}$ & $\chi^2_{tabulated}$ values

a) If $\chi^2_{cal} < \chi^2_{tab}$, then we accept null hypothesis

b) If $\chi^2_{cal} > \chi^2_{tab}$, then we reject null hypothesis

7/12/21

① On the basis of information given below about the treatment of 200 patients suffering from a disease, state whether the new treatment is comparatively superior to the conventional treatment

	Favorable	Not favorable	Total
New	60	30	90
Conventional	40	70	110
	100	100	

Sol:

Step-1 → Set up null hypothesis i.e H_0: There is no significant difference between new & conventional treatment

Step-2→ Set up alternative hypothesis H_1: There is a significant difference between new & conventional treatmen

Step-3→ Take $\alpha = 0.05$

Step-4→ Calculate the test statistic i.e

$$\chi^2_{cal} = \frac{(O_i - E_i)^2}{E_i}$$

To find expected frequencies

E_i

$\frac{\cancel{100} \times 90}{\cancel{200}\ 2} = 45$	$\frac{\cancel{100} \times 90}{\cancel{200}\ 2} = 45$
$\frac{\cancel{100} \times 110}{\cancel{200}\ 2} = 55$	$\frac{\cancel{100} \times 110}{\cancel{200}\ 2} = 55$

O_i	E_i	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
60	45	$(60-45)^2 = 225$	$\frac{225}{45} = 5$
30	45	$(30-45)^2 = 225$	$\frac{225}{45} = 5$
40	55	$(40-55)^2 = 225$	$\frac{225}{55} = 4.09$
70	55	$(70-55)^2 = 225$	$\frac{225}{55} = 4.09$
			18.18

$$\therefore \chi^2_{cal} = 18.18$$

Step-5→

$DoF = (\text{no. of rows} - 1)(\text{no. of columns} - 1)$ $\qquad \alpha = 0.05$

$= (2-1)(2-1)$

$v = 1$

$$\chi^2_{tab} = 3.841$$

$$\chi^2_{cal} > \chi^2_{tab}$$

$\therefore$ We reject null hypothesis

② Given, the following table for hair colour & eye colour. Find the value of χ^2. Is there good assosciation between the two.

Hair Colour

Eye Colour		Fair	Brown	Black	Total
	Blue	15	5	20	40
	Grey	20	10	20	50
	Brown	25	15	20	60
	Total	60	30	60	150

Sol: Step-1 → H_0: There is no significant difference between the two

Step-2 → H_1: There is a significant difference between the two

Step-3 → $\alpha = 0.05$

Step-4 →

$$\chi^2_{cal} = \frac{(O_i - E_i)^2}{E_i}$$

E_i

$\frac{\cancel{60}^{2}\times 40}{\cancel{150}\,5} = 16$	$\frac{\cancel{30}^{1}\times 40}{\cancel{150}\,5} = 8$	$\frac{60\times 40}{150} = 16$
$\frac{60\times 50}{150} = 20$	$\frac{30\times 50}{150} = 10$	$\frac{60\times 50}{150} = 20$
$\frac{{}^{2}\cancel{60}\times 60}{\cancel{150}\,5} = 24$	$\frac{60\times \cancel{30}^{1}}{\cancel{150}\,5} = 12$	$\frac{{}^{2}\cancel{60}\times 60}{\cancel{150}\,5} = 24$

O_i	E_i	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
15	16	$(15-16)^2 = 1$	$\frac{1}{16} = 0.06$
5	8	$(5-8)^2 = 9$	$\frac{9}{8} = 1.12$
20	16	$(20-16)^2 = 16$	$\frac{16}{16} = 1$
20	20	$(20-20)^2 = 0$	0
10	10	$(10-1)^2$	0
20	20	$(20-20)^2 = 0$	0
25	24	$(24-25)^2 = 1$	$\frac{1}{24} = 0.04$
15	12	$(15-12)^2 = 9$	$\frac{9}{12} = 0.75$
20	24	$(20-24)^2 = 16$	$\frac{16}{24} = 0.67$
			3.64

$\chi^2_{cal} = 3.64$

Step -5 →

$\alpha = 0.05$

$DOF = (3-1)(3-1)$

$= 2\times 2$

$= 4$

$\chi^2_{tab} = 9.488$

$\chi^2_{cal} < \chi^2_{tab}$

∴ We accept null hypothesis

③ From the following data find whether there is any significant like in the habit of taking soft drinks among the categories of employees. Use χ^2-distribution with $LOS = 0.05$

Sol:

Employees

Soft Drinks	Clerks	Teachers	officers	
Pepsi	10	25	65	100
Thumsup	15	30	65	110
Fanta	50	60	30	140
	75	115	160	350

Sol:

Step-1 → H_0: There is no significant difference in the habit of taking soft drinks

Step-2 → H_1: There is a significant difference in the habit of taking soft drinks.

Step-3 → $\alpha = 0.05$

Step-4 →

$$\chi^2_{cal} = \frac{(O_i - E_i)^2}{E_i}$$

E_i

$\frac{75\times100}{350} = 21.42$	$\frac{115\times100}{350} = 32.85$	$\frac{160\times100}{350} = 45.71$
$\frac{75\times110}{350} = 23.57$	$\frac{115\times110}{350} = 36.14$	$\frac{160\times110}{350} = 50.28$
$\frac{75\times140}{350} = 30$	$\frac{115\times140}{350} = 46$	$\frac{160\times140}{350} = 64$

O_i	E_i	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
10	21.42	$(10-21.42)^2 = 130.41$	$\frac{130.41}{21.42} = 6.09$
25	32.85	$(25-32.85)^2 = 61.62$	$\frac{61.62}{32.85} = 1.87$
65	45.71	$(65-45.71)^2 = 372.10$	$= \frac{372.10}{45.71} = 8.14$
15	23.57	$(15-23.57)^2 = 73.44$	$\frac{73.44}{23.57} = 3.11$
30	36.14	$(30-36.14)^2 = 37.69$	$\frac{37.69}{36.14} = 1.04$
65	50.28	$(65-50.28)^2 = 216.67$	$\frac{216.67}{50.28} = 4.30$
50	30	$(50-30)^2 = 400$	$\frac{400}{30} = 13.33$
60	46	$(60-46)^2 = 196$	$\frac{196}{46} = 4.26$
30	64	$(30-64)^2 = 1156$	$\frac{1156}{64} = 18.06$

$= 60.2$

$\chi^2_{cal} = 60.2$

Step-5→

$V = (3-1)(3-1)$ $\alpha = 0.05$

$= 4$

$\chi^2_{tab} = 9.488$

$\chi^2_{cal} > \chi^2_{tab}$

∴ We reject null hypothesis

(4) Two researchers adopted sampling techniques while investigating some group of students to find the no. of students falling into different intelligent levels. The result is as follows.

Researchers	Below Avg	Avg	Above Avg	Genius	Total
X	86	60	44	10	200
Y	40	33	25	2	100
Total	126	93	69	12	300

Using χ^2-distribution Test the hypothesis at 5% Los

ol:

tep-1→H_0: There is no significant difference between e intelligent levels.

Step-2 → H_1: There is a significant difference between the intelligent levels.

step-3 → $\alpha = 0.05$

Step-4 →

$$\chi^2_{cal} = \frac{(O_i - E_i)^2}{E_i}$$

E_i

$\frac{126 \times 200}{300} = 84$	$\frac{93 \times 200}{300} = 62$	$\frac{69 \times 200}{300} = 46$	$\frac{12 \times 200}{300} = 8$
$\frac{126 \times 100}{300} = 42$	$\frac{93 \times 100}{300} = 31$	$\frac{69 \times 100}{300} = 23$	$\frac{12 \times 100}{300} = 4$

O_i	E_i	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
86	84	$(86-84)^2 = 4$	$\frac{4}{84} = 0.04$
60	62	$(60-62)^2 = 4$	$\frac{4}{62} = 0.06$
44	46	$(44-46)^2 = 4$	$\frac{4}{46} = 0.08$
10	8	$(10-8)^2 = 4$	$\frac{4}{8} = 0.5$
40	42	$(40-42)^2 = 4$	$\frac{4}{42} = 0.09$
33	31	$(33-31)^2 = 4$	$\frac{4}{31} = 0.12$
25	23	$(25-23)^2 = 4$	$\frac{4}{23} = 0.17$
2	4	$(2-4)^2 = 4$	$\frac{4}{4} = 1$
			2.06

$$\chi^2_{cal} = 2.06$$

Step - 5 →

$\alpha = 0.05$

$V = (4-1)(2-1)$

$= 3$

$\chi^2_{tab} = 7.815$

$\chi^2_{cal} < \chi^2_{tab}$

$\therefore$ We accept null hypothesis

$\boxed{8/12/21}$

Correlation: Correlation is a statistical analysis which gives relationship between two or more Variables

Eg: The relation between the price & production of a commodity

Types of Correlation:

1) Positive & Negative Correlation
2) Simple & Multiple Correlation
3) Partial & Total Correlation
4) Linear & Non-linear Correlation

Method – I Karl Pearson Correlation Coefficient :

Karl Pearson Correlation Coefficient is a statistical technique used for analyzing the behaviour of two or more variables. It is denoted by "r". This method is also known as Product – Moment Correlation coefficient

* The formula for correlation coefficient is

$$r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}}$$

where $X = (x - \bar{x})$
$Y = (y - \bar{y})$.

NOTE :

① Always the correlation coefficient limits are

$-1 \leq r \leq 1$

② If $r = 1$ the correlation is perfect & positive.

③ If $r = -1$ the correlation is perfect & negative.

④ If $r = 0$ there is no relationship between the variables.

① Find is there any significant correlation between the heights & weights given below.

Height (inches)	57	59	62	63	64	65	55	58	57
Weight	113	117	126	126	130	129	111	116	112

Sol: The correlation coefficient is

$$\gamma = \frac{\sum XY}{\sum X^2 \sum Y^2}$$

$X = x - \bar{x}$

$Y = y - \bar{y}$

$$\bar{x} = \frac{57+59+62+63+64+65+55+58+57}{9}$$

$= \frac{540}{9}$

$= 60$

$$\bar{y} = \frac{113+117+126+126+130+129+111+116+112}{9}$$

$= \frac{1080}{9}$

$= 120$

x	y	$x-\bar{x}$	$y-\bar{y}$	X^2	Y^2	XY
57	113	57−60 = −3	113−120 = −7	9	49	21
59	117	59−60 = −1	117−120 = −3	1	9	3
62	126	62−60 = 2	126−120 = 6	4	36	12
63	126	63−60 = 3	126−120 = 6	9	36	18
64	130	64−60 = 4	130−120 = 10	16	100	40
65	129	65−60 = 5	129−120 = 9	25	81	45
55	111	55−60 = −5	111−120 = −9	25	81	45
58	116	58−60 = −2	116−120 = −4	4	16	8
57	112	57−60 = −3	112−120 = −8	9	64	24

$\sum x^2 = 102$ $\sum xy = 216$

$\sum y^2 = 472$

$$\gamma = \frac{216}{\sqrt{102 \times 472}}$$

$$\gamma = 0.98$$

Since γ is positive

$\therefore$ The relation between height & weight are Positively correlated.

② Find Karl Pearson coefficient of correlation for the following data

a)

Wages	100	101	102	102	100	99	97	98	96	95
Cost of Living	98	99	99	97	95	92	95	94	90	91

b)

x	28	41	40	38	35	33	40	32	36	33
y	23	34	33	34	30	26	28	31	36	38

Sol:

$$\bar{x} = \frac{100+101+102+102+100+99+97+98+96+95}{100}$$

$$= \frac{990}{100}$$

$$\bar{x} = 99$$

$$\bar{y} = \frac{98+99+99+97+95+92+95+94+90+91}{10}$$

$$= \frac{950}{10}$$

$$\bar{y} = 95$$

x	y	$x-\bar{x}$	$y-\bar{y}$	X^2	Y^2	ΣXY
100	98	$100-99$ $=1$	$98-95$ $=3$	1	9	3
101	99	$101-99$ $=2$	$99-95$ $=4$	4	16	8
102	99	$102-99$ $=3$	$99-95$ $=4$	9	16	12
102	97	$102-99$ $=3$	$97-95$ $=2$	9	4	6
100	95	$100-99$ $=1$	$95-95$ $=0$	1	0	0
99	92	$99-99$ $=0$	$92-95$ $=-3$	0	9	0
97	95	$97-99$ $=-2$	$95-95$ $=0$	4	0	0
98	94	$98-99$ $=-1$	$94-95$ $=-1$	1	1	1
96	90	$96-99$ $=-3$	$90-95$ $=-5$	9	25	15
95	91	$95-99$ $=-4$	$91-95$ $=-4$	16	16	16

$\Sigma X^2 = 54$ $\Sigma XY = 61$

$\Sigma Y^2 = 96$

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$

$$= \frac{61}{\sqrt{54 \times 96}}$$

$$= 0.8472$$

Since r is positive

$\therefore$ The relation between them is positively correlated b)

$$\bar{x} = \frac{28+41+40+38+35+33+40+32+36+33}{10}$$

$$= \frac{356}{10}$$

$$= 35.6$$

$$\bar{y} = \frac{23+34+33+34+30+26+28+31+36+38}{10}$$

$$= \frac{313}{10}$$

$$= 31.3$$

$\bar{x}$	y	$x-\bar{x}$	$y-\bar{y}$	x^2	y^2	Σxy
28	23	28−35.6 = −7.6	23−31.3 = −8.3	57.76	68.89	63.08
41	34	41−35.6 = 5.4	34−31.3 = 2.7	29.16	7.29	14.58
40	33	40−35.6 = 4.4	33−31.3 = 1.7	19.36	2.89	7.48
38	34	38−35.6 = 2.4	34−31.3 = 2.7	5.76	7.29	6.48
35	30	35−35.6 = −0.6	30−31.3 = −1.3	0.36	1.69	0.78
33	26	33−35.6 = −2.6	26−31.3 = −5.3	6.76	28.09	13.78
40	28	40−35.6 = 4.4	28−31.3 = −3.3	19.36	10.89	−14.52
32	31	32−35.6 = −3.6	31−31.3 = −0.3	12.96	0.09	1.08
36	36	36−35.6 = 0.4	36−31.3 = 4.7	0.16	22.09	1.88
33	38	33−35.6 = −2.6	38−31.3 = 6.7	6.76	44.89	−17.42

$\Sigma x^2 = 158.4$

$\Sigma xy = 77.2$

$\Sigma y^2 = 194.1$

$$\gamma = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

$$= \frac{77.2}{\sqrt{158.4 \times 194.1}}$$

$$= 0.44$$

Since γ is positive,

$\therefore$ The relation between x & y are positively correlated.

Spearman's Rank Correlation Coefficient:

* This method is used for finding the coefficient correlation by ranks. This method is based on rank, is useful in dealing with qualitative characteristics such as character, intelligence, beauty etc. It is based on the rank given to the observations. The formula for Spearman's rank correlation coefficient is given by

$$\rho = 1 - \frac{6\sum D^2}{N(N^2-1)}$$

There are 2 types of problems for finding rank correlation coefficient

When the ranks are given (not repeated) the formula this type is

$$\rho = 1 - \frac{6\sum D^2}{N(N^2-1)}$$

where $D =$ first rank $-$ second rank

$D = R_1 - R_2$ (or) $x - y$

$N =$ no. of observations

2) When the ranks are not given and repeated then the formula is

$$\rho = 1 - 6\left\{\frac{\sum D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \cdots\cdots}{N(N^2-1)}\right\}$$

where $m_1, m_2, m_3 \ldots$ are the ranks which are repeated

① Following are the ranks obtained by 10 students in 2 subjects statistics & Mathematics. To what extent the knowledge of the students in 2 subjects is related.

Statistics	1	2	3	4	5	6	7	8	9	10
Mathematics	2	4	1	5	3	9	7	10	6	8

Sol: In this problem, the ranks are given and they are not repeated

$\therefore$ The rank correlated coefficient is

$$\rho = 1 - \frac{6\sum D^2}{N(N^2-1)}$$

Statistics x (or) R_1	Mathematics y (or) R_2	$D = x-y$ (or) $R_1 - R_2$	D^2
1	2	$1-2=-1$	1
2	4	$2-4=-2$	4
3	1	$3-1=2$	4
4	5	$4-5=-1$	1
5	3	$5-3=2$	4
6	9	$6-9=-3$	9
7	7	$7-7=0$	0
8	10	$8-10=-2$	4
9	6	$9-6=3$	9
10	8	$10-8=2$	4

$\Sigma D^2 = 40$

$$\rho = 1 - \frac{6 \Sigma D^2}{N(N^2-1)}$$

$$= 1 - \frac{6(4\cancel{0})}{1\cancel{0}(100-1)}$$

$$= 1 - \frac{24}{99}$$

$$= \frac{75}{99}$$

$$= 0.76$$

(2) Ten competitors in a musical test were ranked by three judges A, B & C in following order.

Ranks by A	1	6	5	10	3	2	4	9	7	8
Ranks by B	3	5	8	4	7	10	2	1	6	9
Ranks by C	6	4	9	8	1	2	3	10	5	7

...ng rank correlation method, discuss which pair ...udges has the nearest approach to common likings in music.

Sol: Here $N=10$

Ranks by A (x)	Ranks by B (y)	Ranks by C (z)	$D_1 = x-y$	$D_2 = y-z$	$D_3 = x-z$	D_1^2	D_2^2	D_3^2
1	3	6	-2	-3	-5	4	9	25
6	5	4	1	1	2	1	1	4
5.	8	9	-3	-1	-4	9	1	16
10	4	8	6	-4	2	36	16	4
3	7	1	-4	6	2	16	36	4
2	10	2	-8	8	0	64	64	0
4	2	3	2	-1	1	4	1	1
9	1	10	8	-9	-1	64	81	1
7	6	5	1	1	2	1	1	4
8	9	7	-1	2	1	1	4	1
						200	214	60

$$\rho(x,y) = 1 - \frac{6\sum D_1^2}{N(N^2-1)} = 1 - \frac{6(200)}{10(100-1)}$$

$$= 1 - \frac{\cancel{120}^{40}}{\cancel{99}_{33}}$$

$$= \frac{-7}{33}$$

$$= -0.212$$

$$\rho(y,z) = 1 - \frac{6\sum D_2^2}{N(N^2-1)} = 1 - \frac{6(214)}{10(100-1)}$$

$$= 1 - \frac{1284}{990}$$

$$= 1 - 1.296$$

$$= -0.296$$

$$\rho(x,z) = 1 - \frac{6\sum D_3^2}{N(N^2-1)}$$

$$= 1 - \frac{6(60)}{10(100-1)}$$

$$= 1 - \frac{\cancel{36}^{\cancel{12}\,4}}{\cancel{99}_{\cancel{33}\,11}}$$

$$= \frac{7}{11}$$

$$= 0.636$$

Here judges A and C are positively correlated. Hence A and C have the common likings in the music.

③ A random sample of 5 college students are selected and their grades in Mathematics & Statistics are found to be

Mathematics	85	60	73	43	90
Statistics	93	75	65	50	80

Sol: In this problem only gradings are given, we have to assign the ranks for the given data.

Mathematics	Statistics	Rank (x)	Rank (y)	$D=x-y$	D^2
85	93	2	1	1	1
60	75	4	3	1	1
73	65	3	4	-1	1
43	50	5	5	0	0
90	80	1	2	-1	1

$\sum D^2 = 4$

$$\rho = 1 - \frac{6\sum D^2}{N(N^2-1)}$$

$$= 1 - \frac{6(4)}{5(25-1)}$$

$$\rho = 1 - \frac{\cancel{24}}{5(\cancel{24})}$$

$$= 1 - \frac{1}{5}$$

$$= \frac{4}{5}$$

$$= 0.8$$

(4) The ranks of 16 students in Mathematics & Statistics are as follows (1,1) (2,10) (3,3) (4,4) (5,5) (6,7) (7,2) (8,6) (9,8) (10,11) (11,15) (12,9) (13,14) (14,12) (15,16) (16,13)

Sol:

Mathematics (x)	Statistics (y)	$D = x - y$	D^2
1	1	0	0
2	10	-8	64
3	3	0	0
4	4	0	0
5	5	0	0
6	7	-1	1
7	2	5	25
8	6	2	4
9	8	1	1
10	11	-1	1
11	15	-4	16
12	9	3	9
13	14	-1	1
14	12	2	4
15	16	-1	1
16	13	3	9

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

$$= 1 - \frac{6(136)}{16(256-1)}$$

$$= 1 - \frac{816}{4080}$$

$$= 1 - 0.2$$

$$\rho = 0.8$$

⑤ From the following data calculate the rank correlation coefficient

X	48	33	40	9	16	16	65	24	16	57
Y	13	13	24	6	15	4	20	9	6	19

Sol: From the given data we observed that the ranks are repeated, the rank correlation coefficient is

$$\rho = 1 - 6\left\{\frac{\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \cdots}{N(N^2-1)}\right\}$$

$\frac{6+5}{2} = 5.5$

$\frac{8+9}{2} = 8.5$

$N = 10$

x	y	Rank (x)	Rank (y)	$D = x - y$	D^2
48	13	3	5.5	-2.5	6.25
33	13	5	5.5	-0.5	0.25
40	24	4	1	3	9
9	6	10	8.5	1.5	2.25
16	15	8	4	4	16
16	4	8	10	-2	4
65	20	1	2	-1	1
24	9	6	7	-1	1
6	6	8	8.5	-0.5	0.25
7	19	2	3	-1	1

$\Sigma D^2 = 41$

In "x", 16 is repeated 3 times $m_1 = 3$

In "y", 13 is repeated 2 times $m_2 = 2$

In "y", 6 is repeated 2 times $m_3 = 2$

$$\rho = 1 - 6\left[\frac{41 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)}{10(100-1)}\right]$$

$$= 1 - 6\left[\frac{41 + \frac{24+6+6}{12}}{10(99)}\right]$$

$$= 1 - 6\left[\frac{41+3}{990}\right]$$

$$= 1 - \frac{264}{990}$$

$$\rho = 0.733$$

⑥ Obtain the rank correlation coefficient for the following data.

i)

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

ii) The following table gives the marks obtained by 11 students in English & Telugu. Find the rank correlation coefficient

English	40	46	54	60	70	80	82	85	85	90	95
Telugu	45	45	50	43	40	75	55	72	65	42	70

9)

Sol:

X	Y	Rank (x)	Rank (y)	$D = x - y$	D^2
68	62	4	5	-1	1
64	58	6	7	-1	1
75	68	2.5	3.5	-1	1
50	45	9	10	-1	1
64	81	6	1	5	25
80	60	1	6	-5	25
75	68	2.5	3.5	-1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16

$\Sigma D^2 = 72$

$m_1 = 2$ $\quad N = 10$

$m_2 = 3$

$m_3 = 2$

$$\rho = 1 - 6\left[\frac{\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \cdots\cdots +}{N(N^2-1)}\right]$$

$$= 1 - 6\left[\frac{72 + \frac{1}{12}(2^3-2) + \frac{1}{12}(3^3-3) + \frac{1}{12}(2^3-2)}{10(100-1)}\right]$$

$$= 1 - 6\left[\frac{72 + 2 + \frac{1}{2} + \frac{1}{2}}{990}\right]$$

$$1 - \frac{450}{990}$$

$= 0.545$

ii)

Sol:

English	Telugu	Rank (x)	Rank (y)	$D = x - y$	D^2
40	45	11	7.5	3.5	12.25
46	45	10	7.5	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	25
82	55	5	5	0	
85	72	3.5	2	1.5	2.25
85	65	3.5	4	-0.5	0.25
90	42	2	10	-8	64
95	70	1	3	-2	4

$\Sigma D^2 = 140$

$N = 11$

$m_1 = 2$

$m_2 = 2$

$$\rho = 1 - 6\left\{\frac{\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \cdots}{N(N^2-1)}\right\}$$

$$= 1 - 6\left[\frac{140 + \frac{1}{12}(2^3-2) + \frac{1}{12}(2^3-2)}{11(121-1)}\right]$$

$$= 1 - \cancel{6}^{1}\left[\frac{140 + \frac{1}{2} + \frac{1}{2}}{\cancel{120}_{20} \times 11}\right]$$

$$= 1 - \frac{141}{220}$$

$\rho = 0.359$

15/12/21

Regression: The statistical method which helps us to find the value of unknown variable from the known variable is called Regression.

* The line described in the average relationship between the two variables is called Line of Regression

* There are two types of lines of regression

a) The equation of line of regression of X on Y is given by

$$x-\bar{x} = b_{xy}(y-\bar{y}) \quad \text{where} \quad b_{xy} = \frac{\sum XY}{\sum Y^2}$$

$$X = x-\bar{x}$$
$$Y = y-\bar{y}$$

b) The equation of line of regression of Y on X is given by

$$y-\bar{y} = b_{yx}(x-\bar{x}) \qquad b_{yx} = \frac{\sum XY}{\sum X^2}$$

$$X = x-\bar{x}$$
$$Y = y-\bar{y}$$

Uses:

) It is used to estimate the relation between two onomic variables like income & expenditure

It is highly valuable tool in economics & business

is widely used for prediction purpose.

4) We can calculate correlation coefficient with the help of regression coefficient.

5) It is useful in statistical estimation of demand curves, supply curves, production function, cost function etc

① The price indices of cotton & wool are given below for 12 months of a year, obtain the equations of lines of regression.

Cotton	78	77	85	88	87	82	81	77	76	83	97	93
Wool	84	82	82	85	89	90	88	92	83	89	98	99

Sol:

$$\bar{x} = \frac{78+77+85+88+87+82+81+77+76+83+97+93}{12}$$

$$\bar{x} = \frac{1004}{12} \Rightarrow \bar{x} = 83.67 = 84$$

$$\bar{y} = \frac{84+82+82+85+89+90+88+92+83+89+98+99}{12}$$

$$\bar{y} = \frac{1061}{12} \Rightarrow \bar{y} = 88.4 = 88$$

a) The eq^n of line of regression of X on Y is

$$x-\bar{x} = b_{xy}(y-\bar{y})$$

$$b_{xy} = \frac{\Sigma XY}{\Sigma Y^2}$$

$$X = x-\bar{x}$$

$$Y = y-\bar{y}$$

b) The eq^n of line of regression of Y on X is

$$y-\bar{y} = b_{xy}(x-\bar{x})$$

$$b_{xy} = \frac{\Sigma XY}{\Sigma X^2}$$

$$X = x-\bar{x}$$

$$Y = y-\bar{y}$$

$\bar{x}=84 \quad \bar{y}=88$

x	y	$X=x-\bar{x}$	$Y=y-\bar{y}$	XY	X^2	Y^2
78	84	-6	-4	24	36	16
77	82	-7	-6	42	49	36
85	82	1	-6	-6	1	36
88	85	4	-3	-12	16	9
87	89	3	1	3	9	1
82	90	-2	2	-4	4	4
81	88	-3	0	0	9	0
77	92	-7	4	-28	49	16
76	83	-8	-5	40	64	25
83	89	-1	1	-1	1	1
97	98	13	10	130	169	100
93	99	9	11	99	81	121

$\Sigma XY = 287$

$\Sigma X^2 = 488 \quad \Sigma Y^2 = 365$

a) X on Y

$\Rightarrow x-\bar{x} = b_{xy}(y-\bar{y})$

$$b_{xy} = \frac{\Sigma XY}{\Sigma Y^2} = \frac{287}{365} = 0.786$$

$x-84 = 0.786(y-88)$

$x-84 = 0.786y - 69.168$

$x = 0.786y + 84 - 69.168$

$x = 0.786y + 14.832$

b) Y on X

$y-\bar{y} = (x-\bar{x})b_{yx}$

$b_{yx} = \frac{\Sigma XY}{\Sigma X^2}$

$= \frac{287}{488}$

$= 0.588$

$y-88 = 0.588(x-84)$

$y-88 = 0.588x - 49.392$

$y = 0.588x + 88 - 49.392$

$y = 0.588x + 38.608$

② The heights of mothers & daughters are given below Find the equation of lines of regression

Mother	62	63	64	64	65	66	68	70
Daughter	64	65	61	69	67	68	71	65

Sol:

$\bar{x} = \frac{62+63+64+64+65+66+68+70}{8}$

$= \frac{522}{8}$

$= 65.25 = 65$

$\bar{y} = \frac{64+65+61+69+67+68+71+65}{8}$

$= \frac{530}{8}$

$\bar{y} = 66.25$

$= 66$

$\bar{x} = 65 \qquad \bar{y} = 66$

x	y	$X = x - \bar{x}$	$Y = y - \bar{y}$	XY	X^2	Y^2
62	64	-3	-2	6	9	4
63	65	-2	-1	2	4	1
64	61	-1	-5	5	1	25
64	69	-1	3	-3	1	9
65	67	0	1	0	0	1
66	68	1	2	2	1	4
68	71	3	5	15	9	25
70	65	5	-1	-5	25	1

$\Sigma XY = 22 \qquad \Sigma X^2 = 50 \qquad \Sigma Y^2 = 70$

a) The line of regression eq^n of X on Y is

$$x - \bar{x} = b_{xy}(y - \bar{y}) \qquad b_{xy} = \frac{\Sigma XY}{\Sigma Y^2}$$

$$x - 65 = 0.314(y - 66) \qquad = \frac{22}{70}$$

$$x - 65 = 0.314y - 20.724 \qquad = 0.314$$

$$x = 0.314y + 44.276$$

b) The line of regression eq^n of Y on X is

$$y - \bar{y} = b_{yx}(x - \bar{x}) \qquad b_{yx} = \frac{\Sigma XY}{\Sigma X^2}$$

$$y - 66 = 0.44(x - 65) \qquad = \frac{22}{50}$$

$$y = 0.44x + 66 - 28.6 \qquad = 0.44$$

$$y = 0.44x + 37.4$$

3) For the following data

X	1	5	3	2	1	1	7	3
Y	6	1	0	0	1	2	1	5

i) Find the regression line of Y on X & hence find Y if X = 10

ii) Find the regression line of X on Y & hence find X if Y = 2.5

Sol:

$$\bar{x} = \frac{1+5+3+2+1+1+7+3}{8}$$

$$= \frac{23}{8}$$

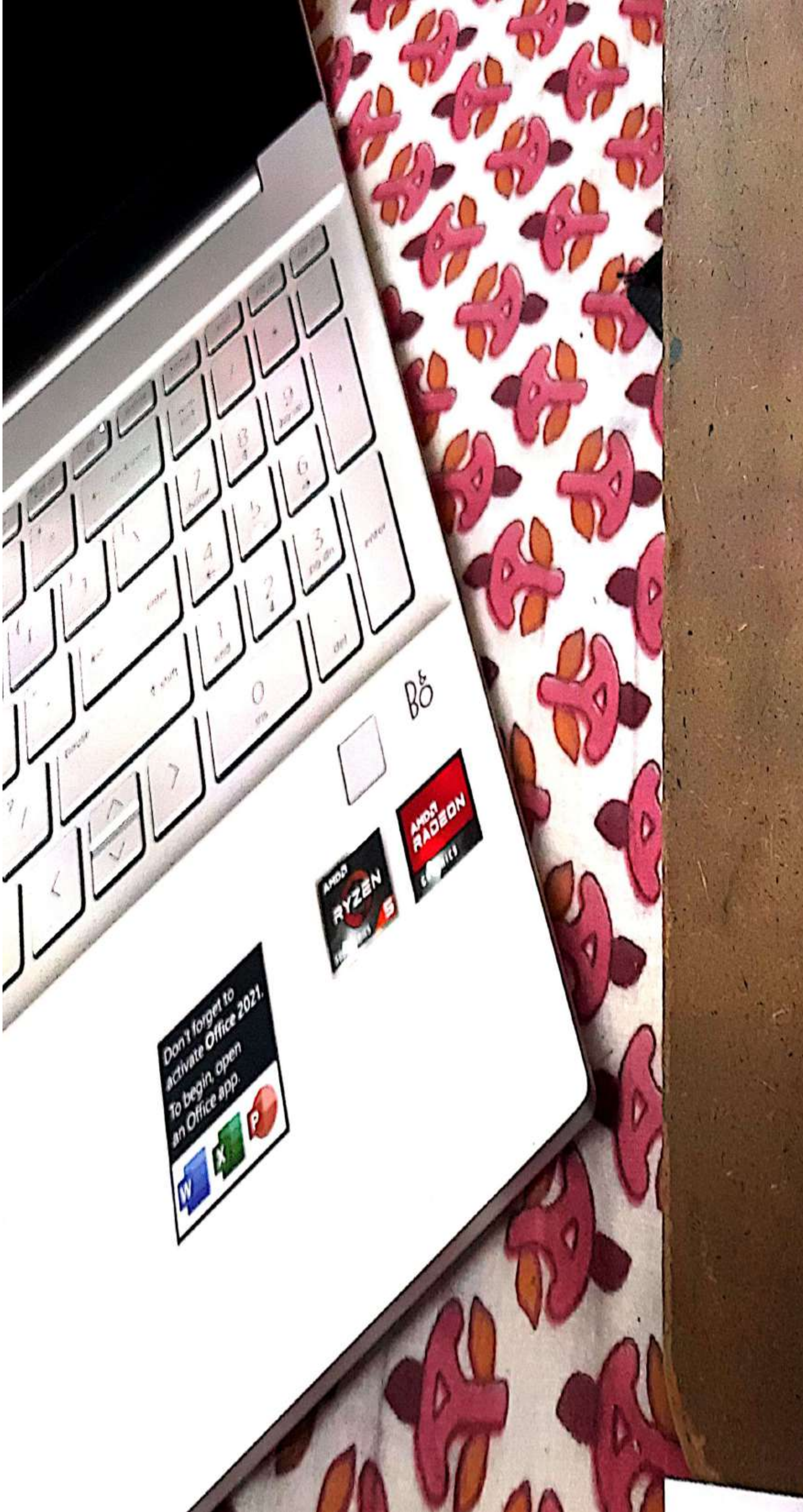
$$\bar{x} = 2.8 \Rightarrow 3 = \bar{x}$$

$$\bar{y} = \frac{6+1+0+0+1+2+1+5}{8}$$

$$= \frac{16}{8} \Rightarrow \bar{y} = 2$$

x	y	$X = x-\bar{x}$	$Y = y-\bar{y}$	XY	X^2	Y^2
1	6	-2	4	-8	4	16
5	1	2	-1	-2	4	1
3	0	0	-2	0	0	4
2	0	-1	-2	2	1	4
1	1	-2	-1	2	4	1
1	2	-2	0	0	4	0
7	1	4	-1	-4	16	1
3	5	0	3	0	0	9

$\sum XY = -10$ $\quad \sum X^2 = 33$ $\quad \sum Y^2 = 36$


Don't forget to activate Office 2021. To begin, open an Office app.
AMD RADEON
AMD RYZEN

a) The eqⁿ of line of regression of Y on X is

$$y-\bar{y} = b_{yx}(x-\bar{x}) \qquad b_{yx} = \frac{\Sigma xy}{\Sigma x^2}$$

$$y-2 = -0.303(x-3) \qquad = \frac{-10}{33}$$

$$y-2 = -0.303x + 0.909 \qquad = -0.303$$

$$y = 2 + 0.909 - 0.303x$$

$$y = -0.303x + 2.909$$

$$x=10 \Rightarrow y = -0.303(10) + 2.909$$

$$y = -3.03 + 2.909$$

$$y = -0.121$$

b) The eqⁿ of line of regression of X on Y is

$$x-\bar{x} = (y-\bar{y})b_{xy} \qquad b_{xy} = \frac{\Sigma xy}{\Sigma y^2}$$

$$x-3 = -0.277(y-2) \qquad = \frac{-10}{36}$$

$$x = -0.277y + 0.554 + 3 \qquad = -0.277$$

$$x = -0.277y + 3.554$$

$$y=2.5 \Rightarrow x = -0.277(2.5) + 3.554$$

$$x = -0.6925 + 3.554$$

$$x = 2.8615$$

a) The eq^n of line of regression of Y on X is

$$y-\bar{y} = b_{yx}(x-\bar{x})$$

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{-10}{33} = -0.303$$

$$y-2 = -0.303(x-3)$$

$$y-2 = -0.303x + 0.909$$

$$y = 2 + 0.909 - 0.303x$$

$$y = -0.303x + 2.909$$

$x = 10 \Rightarrow y = -0.303(10) + 2.909$

$$y = -3.03 + 2.909$$

$$y = -0.121$$

b) The eq^n of line of regression of X on Y is

$$x-\bar{x} = (y-\bar{y})b_{xy}$$

$$b_{xy} = \frac{\Sigma xy}{\Sigma y^2} = \frac{-10}{36} = -0.277$$

$$x-3 = -0.277(y-2)$$

$$x = -0.277y + 0.554 + 3$$

$$x = -0.277y + 3.554$$

$y = 2.5 \Rightarrow x = -0.277(2.5) + 3.554$

$$x = -0.6925 + 3.554$$

$$x = 2.8615$$

20/12/21

UNIT-V

Stochastic Process & Markov Chains

* Stochastic Process is a Greek word which means random (or) chance

* Stochastic deals with models which involve randomness.

* A random variable is a rule (or function) that assign a real number to every outcome of a random experiment.

While, a random process is a rule (or function) that assign a time function to every outcome of a random experiment.

Stochastic Process:

* A stochastic process or random process is defined as a collection of random variables

$\{X(t_n), n=1,2,3\dots\}$

* The random variable $X(t)$ stands for observation at a time "t". It is usually denoted by $X(t,\lambda)$

States: The values assumed by the random variables are called states. The set of all the possible values of an individual random variable X_n of a stochastic process is called state space.

a) The eq^n of line of regression of Y on X is

$$y-\bar{y} = b_{yx}(x-\bar{x})$$

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{-10}{33} = -0.303$$

$$y-2 = -0.303(x-3)$$

$$y-2 = -0.303x + 0.909$$

$$y = 2 + 0.909 - 0.303x$$

$$y = -0.303x + 2.909$$

$x=10 \Rightarrow y = -0.303(10) + 2.909$

$$y = -3.03 + 2.909$$

$$y = -0.121$$

b) The eq^n of line of regression of X on Y is

$$x-\bar{x} = (y-\bar{y})b_{xy}$$

$$b_{xy} = \frac{\Sigma xy}{\Sigma y^2} = \frac{-10}{36} = -0.277$$

$$x-3 = -0.277(y-2)$$

$$x = -0.277y + 0.554 + 3$$

$$x = -0.277y + 3.554$$

$y=2.5 \Rightarrow x = -0.277(2.5) + 3.554$

$$x = -0.6925 + 3.554$$

$$x = 2.8615$$

20/12/21

UNIT-V

Stochastic Process & Markov Chains

* Stochastic Process is a Greek word which means random (or) chance

* Stochastic deals with models which involve randomness.

* A random variable is a rule (or function) that assign a real number to every outcome of a random experiment.

While, a random process is a rule (or function) that assign a time function to every outcome of a random experiment.

Stochastic Process:

* A stochastic process or random process is defined as a collection of random variables

$\{X(t_n); n = 1, 2, 3 \ldots\}$

* The random variable $X(t)$ stands for observation at a time "t". It is usually denoted by $X(t, \lambda)$

States: The values assumed by the random variables are called states. The set of all the possible values of an individual random variable X_n of a stochastic process is called state space.

* The state space is said to be discrete if it contains a finite set of points otherwise it is called Continuous State Space.

Classification:

* The random process is classified into 2 categories
a) Continuous Random Process
b) Discrete Random Process

Continuous Random Process: A continuous random process is one in which the random variable X is continuous and t can have any value between t_1 and t_2.

* Eg: Thermal noise generated in a network, fluctuations in air temperature & air pressure etc...

Discrete Random Process: If the random variable X can assume only certain specific values when t is continuous, it is called Discrete Random Process.

Eg: The voltage available at the one end of a witch because of random opening & closing of e switch is a discrete random process.

* Random Process can also be classified into

a) <u>Deterministic Random Process</u>: If the future values of any sample function can be predicted from the knowledge of past values, then the random process is called Deterministic Process.

Eg: $X(t) = A\cos(\omega t+\theta)$ consist of a family of sine waves and it is completely specified in terms of random variables A and θ then the process is called Deterministic Random Process.

b) <u>Non-Deterministic Random Process</u>: If the future values of a sample function cannot be predicted from the knowledge of past values, then the random process is called Non-Deterministic Random Process.

<u>Classification of Stochastic Process</u>:

* Stochastic process is a function of sample points and time. The sample points may have discrete or continuous values.

$X(t)$ \ t	Continuous	Discrete
Continuous	Continuous Stochastic Process	Discrete Stochastic Process
Discrete	Discrete Stochastic Process	Continuous Stochastic Process

Markov Process:

* A stochastic process is said to be a Markov process if given the value of $X(t)$, the value of $X(v), v>t$ doesnot depend on the value of $X(u)$ for $u<t$.
i.e the future behaviour of Markov process depends only on the present value and not on the past values is called a Markov Process.

* A stochastic process $X(t)$ is said to be Markovian if

Future Present

$$P\left[X(t_{n+1}) \le x_{n+1} \middle| x(t_n)=x_n, X(t_{n-1}=x_{n-1}, \ldots X(t_0)=t\right]$$

$$= P\left[[X(t_{n+1}) \le x(t_n)] = x_n\right] \quad \text{where } t_0 \le t_1 \le \ldots < t_n < t_{n+1}$$

$x_0, x_1, \ldots x_n$ are called the states of process

Markov Chain:

* A sequence of states $[x_n]$ is a Markov chain if each x_n is a random variable and if

$$P\left[X_{n+1} = x_{n+1} \middle| X_n = x_n \ldots X_0 = x_0\right]$$

the constants $\{x_0, x_1, \ldots x_n\}$ are called states of Markov chain.

Transition Probability Matrix: [TPM]

* The probability of moving from one state to another state (or) remaining in the same state during a time period is called TPM.

i.e

The matrix "P" is called TPM with elements P_{ij} i.e $P = P_{ij}$ is the state space

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots\cdots & P_{1n} \\ P_{21} & P_{22} & \cdots\cdots & P_{2n} \\ \vdots & & & \\ P_{m1} & P_{m2} & \cdots\cdots & P_{mn} \end{bmatrix}$$

$i \to$ rows
$j \to$ columns

21/12/21

Stochastic Matrix: A matrix is said to be stochastic if it satisfies the following 3 conditions:

a) The matrix should be a square matrix

b) Always $P_{ij} \geq 0$

c) $\sum P_{ij} = 1$ (i.e the row sum should be 1)

S

1) Which of the following matrices are Stochastic?

a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

Sol: The given matrix is not a square matrix

$\therefore$ The matrix is not a stochastic matrix

b) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Sol: The given matrix is a square matrix, all the elements are non-negative and the sum of each row is 1

$\therefore$ The matrix is a stochastic matrix

c) $\begin{bmatrix} 0 & 1 \\ \frac{1}{3} & \frac{1}{4} \end{bmatrix}$

Sol: The given matrix is a square matrix, all the elements are non-negative but the sum of each row is not 1

$\therefore$ The matrix is not a stochastic matrix

d) $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$

Sol: The given matrix is a square matrix, all the elements are non-negative and the sum of each row is 1

$\therefore$ The matrix is a stochastic matrix

) $\begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}$

: The given matrix is a square matrix but all the
nents are not positive.

he matrix is not a stochastic process

f) $\begin{bmatrix} 0 & 2 \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}$

Sol: The given matrix is a square matrix, and all the elements are non-negative but the sum of each row is not 1

$\therefore$ The matrix is not a stochastic matrix.

g) $\begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{4}{3} \\ \frac{1}{2} & 1 & \frac{1}{2} \end{bmatrix}$

Sol: The given matrix is not a square matrix

$\therefore$ The matrix is not a stochastic matrix

h) $\begin{bmatrix} \frac{15}{16} & \frac{1}{16} \\ \frac{2}{3} & \frac{4}{3} \end{bmatrix}$

Sol: The given matrix is a square matrix, and all the elements are non-negative, but the sum of each row is not 1

$\therefore$ The matrix is not a stochastic matrix

i) $\begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$

Sol: The given matrix is a square matrix, and all the elements are non-negative and the sum of each row is 1

$\therefore$ The given matrix is a stochastic matrix

② Find the values of x, y, z if

$$\begin{bmatrix} 0 & x & \frac{1}{3} \\ 0 & 0 & y \\ \frac{1}{3} & \frac{1}{4} & z \end{bmatrix} \text{ is a TPM}$$

Sol: By definition of Stochastic Process,

$$\sum P_{ij} = 1$$

$x + \frac{1}{3} = 1 \Rightarrow x = 1 - \frac{1}{3}$

$\Rightarrow x = \frac{2}{3}$

$y + 0 = 1 \Rightarrow y = 1$

$z + \frac{1}{3} + \frac{1}{4} = 1 \Rightarrow z = 1 - \frac{7}{12}$

$\Rightarrow z = \frac{5}{12}$

Regular Matrix: A stochastic matrix P is said to be regular if all the entries of some power P^m are positive.

NOTE: A stochastic matrix P is not regular if '1' occurs in the principal diagonal.

Eg: ① Test the following matrices are regular or not.

a) $$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

Sol: In the above matrix, 1 occurs in the principal diagonal elements

∴ The given matrix is not regular

$$C = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$$

$$C^2 = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

$$C^3 = \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

$$C^4 = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$C^5 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{8} & \frac{1}{2} & \frac{3}{8} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$ Since in C^5 all the elements are positive

$\therefore$ The given matrix is a regular matrix

b) $A = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.5 & 0.5 \\ 0.6 & 0.4 & 0 \end{bmatrix}$

Sol:

$$A = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.5 & 0.5 \\ 0.6 & 0.4 & 0 \end{bmatrix}$$

$$A^2 = \begin{bmatrix} \frac{9}{16} & \frac{5}{16} & \frac{1}{8} \\ \frac{3}{10} & \frac{9}{20} & \frac{1}{4} \\ \frac{9}{20} & \frac{7}{20} & \frac{1}{5} \end{bmatrix}$$

Since all the elements are positive

$\therefore$ The given matrix is regular

d) $$B = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

Sol:- $$B = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

$$B^2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{3}{8} & \frac{3}{8} & \frac{1}{4} \end{bmatrix}$$

$$B^3 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{7}{16} & \frac{7}{16} & \frac{1}{8} \end{bmatrix}$$

$$B^4 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{15}{32} & \frac{15}{32} & \frac{1}{16} \end{bmatrix}$$

ce in B^2, B^3, B^4 the place of zeroes is not

ing, the no.of times we multiply also the zero's

be unchanged.

∴ The given matrix is not regular

e) $D = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & 1 & 0 \end{bmatrix}$

Sol:

$$D = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & 1 & 0 \end{bmatrix}$$

$$D^2 = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{8} & \frac{5}{16} & \frac{9}{16} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

$$D^3 = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{5}{32} & \frac{41}{64} & \frac{13}{64} \\ \frac{1}{8} & \frac{5}{16} & \frac{9}{16} \end{bmatrix}$$

Since all the elements are positive

∴ The given matrix is a regular matrix

f) $B = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & 1 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 1 \end{bmatrix}$

Sol: In the above matrix, 1 occurs in the principal diagonal elements

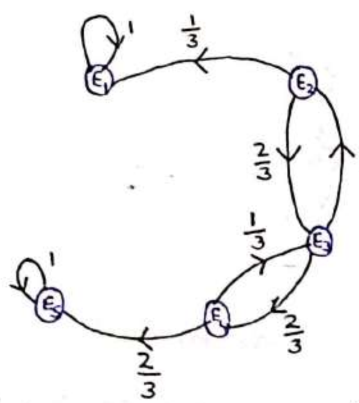
∴ The given matrix is not regular

② Consider the TPM, find its graph

Sol:

$$\begin{array}{cc} & \xleftarrow{\text{States}} \\ & \begin{array}{ccccc} E_1 & E_2 & E_3 & E_4 & E_5 \end{array} \\ \text{States}\left\{\begin{array}{c} E_1 \\ E_2 \\ E_3 \\ E_4 \\ E_5 \end{array}\right. & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Sol:



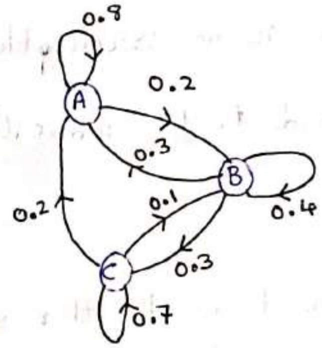
NOTE:

1) "0" indicates no relation between 2 states.

③ Three universities A, B, C are admitting students. :t is given that 80% of the children of A went o A and the rest went to B. 40% of the ildren of B went to B and the rest split ly between A and C. 70% of the children of went to C, and 20% went to A & 10% went to m the Markov Chain & Transition Matrix.

Sol:

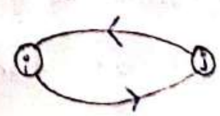
$$\begin{array}{c|ccc} & A & B & C \\ \hline A & & & \\ B & & & \\ C & & & \end{array} \quad \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}$$



Classification of Markov Chain:

Accessible: A state j is said to be accessible from state i if $P_{ij}^{(n)} > 0 \ \forall n$ (or) if there exists a Path from i to j, we say j is accessible from i.

Communicate: Two accessible states i and j are said to communicate if state i is accessible from State j and state j is accessible from state i, & can communicate with itself.



Irreducable markov chain: If $P_{ij}^{(n)} > 0$ for some n and for any i and j then every state can be reached from every other state such a chain is said to be irreducable markov chain. The transition probability matrix of irreducable chain is an irreducable matrix

* Otherwise, the chain is said to be reducible (or) non-irreducable.

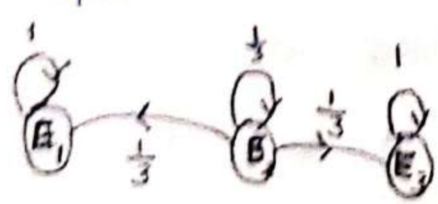
NOTE: All states are reachable from all other states. Then the markov chain is called irreducable.

Non-null Persistant: A finite irreducible markov chain is called Non-null persistant. In the above example, the no. of states are 3

∴ Above Markov chain is finite irreducible Markov chain. Hence it is non-null persistant.

Absorbing State: A state "i" is said to be an absorbing tate if and only if $P_{ii} = 1$ (or) if it has atleast e absorbing state and it is possible to go from ny non-absorbing state to atleast one absorbing te in one or more steps.

$$P = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix}$$



* States with probability=1 are called Absorbing States.

* In above example, $P_{11} = 1 \Rightarrow E_1$ state is Absorbing State

$P_{33} = 1 \Rightarrow E_3$ state is Absorbing state

Return: A state i of a Markov chain is called a return state if $P_{ij}^{(n)} > 0$ for some $n \geq 1$.

Period, Periodic and Aperiodic: The period of a return state is defined as the greatest common divisor of all $m \ni P_{ii}^{(m)} > 0$ i.e $d_i = G.C.D\{m / P_{ii}^{m} > 0\}$

* A state i is said to be periodic if $d_i > 1$ and aperiodic if $d_i = 1$

Transient State: If there exists atleast one path which cannot be returned to the original state, then that state is called transient state.

Recurrent State: If a state is not transient, then it is recurrent.

Ergodic State: If a state is positive recurrent and aperiodic, then it is called ergodic.

Absorbing State: If probability of state is 1 i.e $P_{ii} = 1$ then that state is called Absorbing state

* If a Markov chain contains atleast one absorbing state, then it is called Absorbing Markov chain.

22/12/21

Nature

→ Communicate
+
→ Finite Markov Chain
↓
Irreducible
⇓
→ Non-null persistant

→ Period GCD = 1 → aperiodic
GCD > 1 → Periodic

→ Non-null persistant + aperiodic ⇒ Ergodic

→ Non-null persistant + Periodic ⇒ Not ergodic

→ Absorbing state

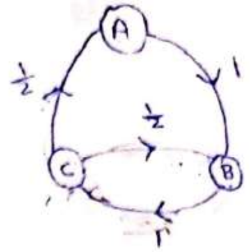
→ $GCD\{2,3,5,7,\ldots\} = 1$

→ $GCD\{2,4,6.\} = 2$

1) 3 boys A, B and C are throwing a ball to each other. always throws the ball to B and B always throws ie ball to C, C is likely to throw the ball to and A. Show that the process is Markov process. d the transition matrix and classify the es. Do all the states are ergodic?

The TPM is

$$\begin{array}{c|ccc} & A & B & C \\ \hline A & 0 & 1 & 0 \\ B & 0 & 0 & 1 \\ C & \frac{1}{2} & \frac{1}{2} & 0 \end{array}$$



State A: A is accessible from B and C

State B: B is accessible from A and C

State C: C is accessible from A and B

$\therefore$ All states are accessible from all other states

$\therefore$ All states are communicate.

Here the no. of states in Markov chain are finite

Since all the states communicate each other

$\therefore$ A, B, C are irreducible

All states are irreducible and it is a finite Markov chain

$\therefore$ All. the states are non-null persistant

To find the period:

Period of A:

$$GCD\{3, 5, \ldots\} = 1$$

Since $GCD = 1$

$\therefore$ Period of $A = 1$

$\therefore$ State A is aperiodic

Period of B:

$GCD\{2,3..\} = 1$

Since $GCD = 1$

$\therefore$ Period of $B = 1$

$\therefore$ State B is aperiodic

Period of C:

$GCD\{2,3,...\} = 1$

Since $GCD = 1$

$\therefore$ Period of $C = 1$

$\therefore$ State C is aperiodic

$\therefore$ Three states A,B,C are aperiodic

Since A,B,C are non-null persistant and aperiodic

$\therefore$ The three states A,B,C are ergodic

The three states are not absorbing states (i.e

No state has 1 in the diagonal elements]

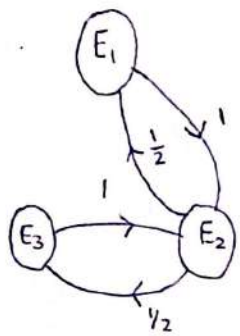
Hence it is not a absorbing Markov Chain.

(2) Find the nature of states of Markov chain with

TPM

$$P = \begin{array}{c|ccc} & E_1 & E_2 & E_3 \\ \hline E_1 & 0 & 1 & 0 \\ E_2 & 1/2 & 0 & 1/2 \\ E_3 & 0 & 1 & 0 \end{array}$$

Sol:



State E_1: E_1 is accessible from E_2 and E_3

State E_2: E_2 is accessible from E_1 and E_3

State E_3: E_3 is accessible from E_1 and E_2

$\therefore$ All the states are accessible from all other states

$\therefore$ All states are communicate

Hence the no. of states in Markov chain are finite

Since all the states communicate each other

$\therefore$ E_1, E_2, E_3 are irreducible

All states are irreducible and it is a finite Markov chain

$\therefore$ All the states are non-null persistant

To find the period:

Period of E_1:

$$GCD\{2,4,\cdots\}=2$$

Since $GCD=2$

Period of $E_1=2$

$\therefore$ State E_1 is periodic

Period of E_2:

$$GCD\{2,2\}=2$$

Since $GCD=2$

Period of $E_2=2$

$\therefore$ State E_2 is periodic

Period of E_3:

$$GCD\{2,4\}=2$$

Since $GCD=2$

Period of $E_3=2$

$\therefore$ State E_3 is periodic

$\therefore$ Three states are Periodic

Since A, B, C are non-null Persistant and periodic

$\therefore$ The three states A, B, C are not ergodic

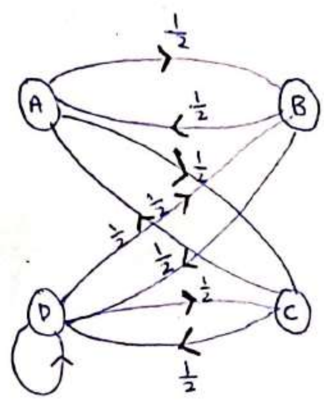
The three states are not absorbing states

ence it is not a absorbing Markov chain

(3) Check whether the following Markov chain is ergodic.

$$P = \begin{array}{c c} & \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \end{array}$$

Sol:



State A: A is accessible from B, C, D

State B: B is accessible from A, C, D

State C: C is accessible from A, B, D

State D: D is accessible from A, B, C

$\therefore$ All the states are accessible from all other states

$\therefore$ All states are communicate

The no. of states in Markov chain are finite

Since all the states communicate with each other

$\therefore$ A, B, C, D are irreducible

All states are irreducible and it is a finite Markov chain

$\therefore$ All the states are non-null persistent

To find the period:

Period of A:

$GCD\{2,4,6..\} = 2$

$GCD = 2$

Period of $A = 2$

$\therefore$ State A is periodic

Period of B:

$GCD\{2,4,6\} = 2$

$GCD = 2$

Period of $B = 2$

$\therefore$ State B is periodic

Period of C:

$GCD\{2,4,6\} = 2$

$GCD = 2$

Period of $C = 2$

$\therefore$ State C is periodic

Period of D:

$GCD\{\cancel{6}\,2,4\} = 2$

$GCD = 2$

Period of $D = 2$

$\therefore$ State D is periodic

Four
∴ ~~Three~~ States are periodic

Since A, B, C are non-null persistant & periodic

∴ The four states A, B, C, D are not ergodic

The four states are not absorbing states (No 1 in P.D)

∴ It is not a absorbing Markov chain

27/12/21

Steady State Condition:

* In many Markov chains the probability for a particular state will approach a limiting value as time tends to infinity. In other words, the far future the probability will not change from one transition to next transition state. These limiting values are called stable probabilities.

* If a system is such that each state has a probability equal to its stable probability, the probability will persist for all the time, then the system is said to be in steady state condition.

Steady State Vector (or) Equilibrium Vector: If a Markov chain with transition matrix 'P' is regular then there exist a unique vector "E" such that for any probability vector 'V' and for large values of 'n' we have $VP^n = V$. Here vector V is called Equilibrium Vector of a Markov Chain. This is also called Long Range Trend of the Markov Chain.

Probability Vector: A probability vector is a matrix of only one row having non-negative elements whose sum is equal to '1'.

NOTE: If a Markov chain with transition matrix 'P' is regular then there exist a probability vector 'V' such that $VP = V$ which gives the Long range trend of the Markov Chain.

① Find the Long Range Trend (or) Steady State Vector (or) Equilibrium vector for the Markov chain with transition matrix.

$$P = \begin{bmatrix} 0.65 & 0.28 & 0.07 \\ 0.15 & 0.67 & 0.18 \\ 0.12 & 0.36 & 0.52 \end{bmatrix}$$

Sol: A vector V is said to be an Equilibrium vector if

$VP = V$ where $V = [v_1 \; v_2 \; v_3]$

$$[v_1 \; v_2 \; v_3] \begin{bmatrix} 0.65 & 0.28 & 0.07 \\ 0.15 & 0.67 & 0.18 \\ 0.12 & 0.36 & 0.52 \end{bmatrix} = [v_1 \; v_2 \; v_3]$$

$0.65V_1 + 0.15V_2 + 0.12V_3 = V_1$

$0.28V_1 + 0.67V_2 + 0.36V_3 = V_2$

$0.07V_1 + 0.18V_2 + 0.52V_3 = V_3$

$-0.35V_1 + 0.15V_2 + 0.12V_3 = 0$ — ①

$-0.33V_2 + 0.28V_1 + 0.36V_3 = 0$ — ②

$0.07V_1 + 0.18V_2 - 0.48V_3 = 0$ — ③

The above equations are linearly dependent

$\therefore$ The sol^n is '0' and we know that the probability never be 0, to avoid this we consider a stable probability vector. From ①, ②, ④ $\quad V_1 + V_2 + V_3 = 1$ ④

$$[V_1 \ V_2 \ V_3] = \begin{bmatrix} \frac{106}{363} & \frac{532}{1089} & \frac{245}{1089} \end{bmatrix}$$

② Find Equilibrium Vector (or) Steady State Vector for the TPM

a) $P = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.1 & 0.4 & 0.5 \\ 0.2 & 0.2 & 0.6 \end{bmatrix}$

Sol: $VP = V$

$$[V_1 \ V_2 \ V_3] \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.1 & 0.4 & 0.5 \\ 0.2 & 0.2 & 0.6 \end{bmatrix} = [V_1 \ V_2 \ V_3]$$

$0.5V_1 + 0.1V_2 + 0.2V_3 = V_1$

$\Rightarrow -0.5V_1 + 0.1V_2 + 0.2V_3 = 0$ ①

$0.2V_1 + 0.4V_2 + 0.2V_3 = V_2$

$\Rightarrow 0.2V_1 - 0.6V_2 + 0.2V_3 = 0$ ②

$0.3V_1 + 0.5V_2 + 0.6V_3 = V_3$

$\Rightarrow 0.3V_1 + 0.5V_2 - 0.4V_3 = 0$ ③

The above equations are linearly dependent

$\therefore$ The sol^n is '0' & we know that the probability never be '0', to avoid this we consider a stable probability vector

$$V_1 + V_2 + V_3 = 1 \quad ④$$

From ①, ② & ④

$$[V_1 \ V_2 \ V_3] = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

b) $P = \begin{bmatrix} 0.25 & 0.75 \\ 0.5 & 0.5 \end{bmatrix}$

Sol: $\begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} 0.25 & 0.75 \\ 0.5 & 0.5 \end{bmatrix} = \begin{bmatrix} V_1 & V_2 \end{bmatrix}$

$0.25V_1 + 0.5V_2 = V_1$

$-0.75V_1 + 0.5V_2 = 0$ — ①

$0.75V_1 + 0.5V_2 = V_2$

$\Rightarrow 0.75V_2 - 0.5V_2 = 0$ — ②

Consider $V_1 + V_2 = 1$ — ③

From ①,②,③ $\begin{bmatrix} V_1 & V_2 \end{bmatrix} = \begin{bmatrix} \frac{2}{5} & \frac{3}{5} \end{bmatrix}$

③ A house wife buys 3 kinds of cereals A, B, C. She never buys the same cereals in successive week. If she buys cereals A, next week she buys B. However if she buys B or C, the next week it is three times likely to buy, A as other cereals. In the long run, how often she buy each of the 3 cereals.

Sol:

$$P = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 3x & 0 & x \\ 3x & x & 0 \end{bmatrix} \end{array}$$

$3x + x = 1$

$x = \frac{1}{4}$

$$P = \begin{bmatrix} 0 & 1 & 0 \\ \frac{3}{4} & 0 & \frac{1}{4} \\ \frac{3}{4} & \frac{1}{4} & 0 \end{bmatrix}$$

$$\begin{bmatrix} V_1 & V_2 & V_3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ \frac{3}{4} & 0 & \frac{1}{4} \\ \frac{3}{4} & \frac{1}{4} & 0 \end{bmatrix} = \begin{bmatrix} V_1 & V_2 & V_3 \end{bmatrix}$$

$$\Rightarrow \frac{3V_2}{4} + \frac{3V_3}{4} = V_1$$

$$\Rightarrow -V_1 + \frac{3V_2}{4} + \frac{3V_3}{4} = 0 \quad \text{—} ①$$

$$V_1 + \frac{V_3}{4} = V_2$$

$$\Rightarrow V_1 - V_2 + \frac{V_3}{4} = 0 \quad \text{—} ②$$

$$\frac{V_2}{4} = V_3$$

$$\Rightarrow \frac{V_2}{4} - V_3 = 0 \quad \text{—} ③$$

Consider $V_1 + V_2 + V_3 = 1$ — ④

From ①,②,④

$$\begin{bmatrix} V_1 & V_2 & V_3 \end{bmatrix} = \begin{bmatrix} \frac{3}{7} & \frac{16}{35} & \frac{4}{35} \end{bmatrix}$$

④ The weather in certain spot is classified as Fair, Cloudy and Rainy. A fair day is followed by a fair day, 60% of the time and by a cloudy day 25% of the time. A Cloudy day is followed by a cloudy day 35% of the time and by a rainy day 25% of the time. A rainy day is followed by a cloudy day 40% of the time & by a rainy day 25% of the time. Initial probabilities are 0.3, 0.3, 0.4

Sol:

$$P = \begin{bmatrix} 0.6 & 0.25 & 0.15 \\ 0.4 & 0.35 & 0.25 \\ 0.35 & 0.4 & 0.25 \end{bmatrix}$$

$$\begin{bmatrix} V_1 & V_2 & V_3 \end{bmatrix} \begin{bmatrix} 0.6 & 0.25 & 0.15 \\ 0.4 & 0.35 & 0.25 \\ 0.35 & 0.4 & 0.25 \end{bmatrix} = \begin{bmatrix} V_1 & V_2 & V_3 \end{bmatrix}$$

$0.6V_1 + 0.4V_2 + 0.35V_3 = V_1$

$\Rightarrow -0.4V_1 + 0.4V_2 + 0.35V_3 = 0$ — ①

$0.25V_1 + 0.35V_2 + 0.4V_3 = V_2$

$\Rightarrow 0.25V_1 - 0.65V_2 + 0.4V_3 = 0$ — ②

$0.15V_1 + 0.25V_2 + 0.25V_3 = V_3$

$\Rightarrow 0.15V_1 + 0.25V_2 - 0.75V_3 = 0$ — ③

Consider $V_1 + V_2 + V_3 = 1$ — ④

From ①, ② & ④

$$\begin{bmatrix} V_1 & V_2 & V_3 \end{bmatrix} = \begin{bmatrix} \frac{155}{318} & \frac{33}{106} & \frac{32}{159} \end{bmatrix}$$

29/12/21

Limiting Probability: If $\{X_n \mid n \geq 0\}$ is a Markov Chain in which all the states are positive, recurrent & aperiodic then the limiting probability is defined as

$$\lim_{n \to \infty} P_{ij}^{(n)} = V_j$$

① Consider a Markov chain with state space $\{1,2,3\}$ and Transition Probability Matrix is

$$P = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \end{array}$$

Then which of the following are true.

a) $\lim_{n\to\infty} P_{11}^{(n)} = \frac{2}{9}$

b) $\lim_{n\to\infty} P_{21}^{(n)} = 0$

c) $Lt_{n\to\infty} P_{32}^{(n)} = \frac{1}{3}$

d) $Lt_{n\to\infty} P_{13}^{(n)} = \frac{1}{3}$

Sol: The given matrix is irreducible, non-null persistent and aperiodic

Let $VP = V$ where V is the statistic vector

$$\begin{bmatrix} V_1 & V_2 & V_3 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} V_1 & V_2 & V_3 \end{bmatrix}$$

$$\frac{1}{2}V_1 + \frac{1}{3}V_3 = V_1$$

$$\Rightarrow -\frac{1}{2}V_1 + \frac{1}{3}V_3 = 0 \quad ①$$

$$\frac{1}{2}V_1 + \frac{1}{2}V_2 + \frac{1}{3}V_3 = V_2$$

$$\Rightarrow \frac{1}{2}V_1 - \frac{1}{2}V_2 + \frac{1}{3}V_3 = 0 \quad ②$$

2

$$\frac{1}{2}V_2 + \frac{1}{3}V_3 = V_3$$

$$\Rightarrow \frac{1}{2}V_2 - \frac{2}{3}V_3 = 0 \quad \text{③}$$

Consider $V_1 + V_2 + V_3 = 1$ ④

From ①, ② & ④

$$\begin{bmatrix} V_1 & V_2 & V_3 \end{bmatrix} = \begin{bmatrix} \frac{2}{9} & \frac{4}{9} & \frac{1}{3} \end{bmatrix}$$

By definition of limiting probability, we have

$$\underset{n\to\infty}{Lt}\, P_{ij}^{(n)} = V_j$$

a) $\underset{n\to\infty}{Lt}\, P_{11}^{(n)} = V_1$

$= \frac{2}{9}$ (True)

b) $\underset{n\to\infty}{Lt}\, P_{21}^{(n)} = V_1$

$= \frac{2}{9}$ (False)

c) $\underset{n\to\infty}{Lt}\, P_{32}^{(n)} = V_2$

$= \frac{4}{9}$ (False)

d) $\underset{n\to\infty}{Lt}\, P_{13}^{(n)} = V_3$

$= \frac{1}{3}$ (True)

Consider Markov chain with state space $S = \{1, 2, 3\}$ &

TPM

$$P = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} \frac{1}{4} & \frac{5}{8} & \frac{1}{8} \\ \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \end{bmatrix} \end{array}$$

Then which of the following are true.

a) $\underset{n\to\infty}{Lt}\ P_{12}^{(n)} = 0$

b) $\underset{n\to\infty}{Lt}\ P_{12}^{(n)} = \underset{n\to\infty}{Lt}\ P_{21}^{(n)}$

c) $\underset{n\to\infty}{Lt}\ P_{22}^{(n)} = 1/8$

d) $\underset{n\to\infty}{Lt}\ P_{21}^{(n)} = \frac{1}{3}$

Sol: Let $VP = V$

$$[V_1\ V_2\ V_3] \begin{bmatrix} \frac{1}{4} & \frac{5}{8} & \frac{1}{8} \\ \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \end{bmatrix} = [V_1\ V_2\ V_3]$$

$\frac{1}{4}V_1 + \frac{1}{4}V_2 + \frac{1}{2}V_3 = V_1$

$\Rightarrow -\frac{3}{4}V_1 + \frac{1}{4}V_2 + \frac{1}{2}V_3 = 0$ — (1)

$\frac{5}{8}V_2 + 0 + \frac{3}{8}V_3 = V_2$

$\Rightarrow \frac{5}{8}V_1 - V_2 + \frac{3}{8}V_3 = 0$ — (2)

$\frac{1}{8}V_1 + \frac{3}{4}V_2 + \frac{1}{8}V_3 = V_3$

$\Rightarrow \frac{1}{8}V_1 + \frac{3}{4}V_2 - \frac{7}{8}V_3 = 0$ — (3)

Consider $V_1 + V_2 + V_3 = 1$ — (4)

From ①,② & ④

$$\begin{bmatrix} V_1 & V_2 & V_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

a) $\underset{n\to\infty}{Lt} P_{12}^{(n)} = V_2$

$= \frac{1}{3}$ (False)

b) $\underset{n\to\infty}{Lt} P_{12}^{(n)} = V_2$

$= \frac{1}{3}$

$\underset{n\to\infty}{Lt} P_{21}^{(n)} = V_1$

$= \frac{1}{3}$

(True)

c) $\underset{n\to\infty}{Lt} P_{22}^{(n)} = V_2$

$= \frac{1}{3}$ (False)

d) $\underset{n\to\infty}{Lt} P_{21}^{(n)} = V_1$

$= \frac{1}{3}$ (True)

***

③ An urn initially contains 5 black balls & 5 white balls. The following experiment is repeated in definitely. A ball is drawn from the urn. If ball is white, it is Put back into the urn, otherwise it is left out. Let X_n be no. of black balls remaining in the urn after "n" draws. Is X_n is a Markov Process. If so,

a) Find appropriate TPM

b) Find one step probability Matrix for P

Find two step probability Matrix for P

What happens to X_n as $n\to\infty$. Write the TPM

... X_n of black balls remaining in the urn.

Here, X_n is the future value depends on the present value X_{n-1}.

$\therefore X_n$ is a Markov process

a)

b)

$$
\begin{array}{cc}
 & \text{Future } X_n \\
\text{Present } X_{n-1} &
\begin{array}{c|cccccc}
 & 0 & 1 & 2 & 3 & 4 & 5 \\
\hline
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & \frac{1}{6} & \frac{5}{6} & 0 & 0 & 0 & 0 \\
2 & 0 & \frac{2}{7} & \frac{5}{7} & 0 & 0 & 0 \\
3 & 0 & 0 & \frac{3}{8} & \frac{5}{8} & 0 & 0 \\
4 & 0 & 0 & 0 & \frac{4}{9} & \frac{5}{9} & 0 \\
5 & 0 & 0 & 0 & 0 & \frac{5}{10} & \frac{5}{10}
\end{array}
\end{array}
$$

$$P_{00} = \{x_n = 0 \mid x_{n-1} = 0\} = \frac{5_{C_1}}{5} = 1$$

P_{01} = Impossible event

$$P_{10} = \{x_n = 0 \mid x_{n-1} = 1\} = \frac{1_{C_1}}{6} = \frac{1}{6}$$

$$P_{11} = \{x_{n-1} = 1 \mid x_{n-1} = 1\} = \frac{5_{C_1}}{6} = \frac{5}{6}$$

c) Two Step Probability Matrix

$$
P^2 = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
\frac{11}{36} & \frac{25}{36} & 0 & 0 & 0 & 0 \\
\frac{1}{21} & \frac{65}{144} & \frac{25}{49} & 0 & 0 & 0 \\
0 & \frac{3}{28} & \frac{225}{448} & \frac{25}{64} & 0 & 0 \\
0 & 0 & \frac{1}{6} & \frac{95}{192} & \frac{25}{81} & 0 \\
0 & 0 & 0 & \frac{2}{9} & \frac{19}{36} & \frac{1}{4}
\end{bmatrix}
$$

d) As $n \to \infty$ Gradually all black balls are removed.

Hence
$$P^n = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(4) A fair die is tossed repeatedly. If X_n denotes the maximum of the numbers occuring in the first n-tosses. Find the TPM. Also find P^2

Sol:

Future

$$P = \text{Present} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{2}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & \frac{3}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & \frac{4}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & \frac{5}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & \frac{6}{6} \end{bmatrix} \end{array}$$

Now, we have to find P^2

$$P^2 = P.P$$

$$P^2 = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{2}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & \frac{3}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & \frac{4}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & \frac{5}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & \frac{6}{6} \end{bmatrix} \times \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{2}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & \frac{3}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & \frac{4}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & \frac{5}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & \frac{6}{6} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{36}+0 & \frac{1}{36}+\frac{2}{36} & \frac{1}{36}+\frac{1}{36}+\frac{3}{36} & \frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{4}{36} & \frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{5}{36} & \frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{6}{36} \\ 0+0 & 0+\frac{4}{36} & 0+\frac{2}{36}+\frac{3}{36} & 0+\frac{2}{36}+\frac{1}{36}+\frac{4}{36} & 0+\frac{2}{36}+\frac{1}{36}+\frac{1}{36}+\frac{5}{36} & 0+\frac{2}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{6}{36} \\ 0 & 0 & \frac{9}{36} & \frac{3}{36}+\frac{4}{36} & \frac{3}{36}+\frac{1}{36}+\frac{5}{36} & 0+0+\frac{3}{36}+\frac{1}{36}+\frac{1}{36}+\frac{6}{36} \\ 0 & 0 & 0 & 0+\frac{16}{36} & \frac{4}{36}+\frac{5}{36} & \frac{4}{36}+\frac{1}{36}+\frac{6}{36} \\ 0 & 0 & 0 & 0 & \frac{25}{36}+0 & \frac{5}{36}+\frac{6}{36} \\ 0 & 0 & 0 & 0 & 0 & \frac{36}{36} \end{bmatrix}$$

$$P^2 = \begin{bmatrix} \frac{1}{36} & \frac{1}{12} & \frac{5}{36} & \frac{7}{36} & \frac{1}{4} & \frac{11}{36} \\ 0 & \frac{1}{9} & \frac{5}{36} & \frac{7}{36} & \frac{1}{4} & \frac{11}{36} \\ 0 & 0 & \frac{1}{4} & \frac{7}{36} & \frac{1}{4} & \frac{11}{36} \\ 0 & 0 & 0 & \frac{4}{9} & \frac{1}{4} & \frac{11}{36} \\ 0 & 0 & 0 & 0 & \frac{25}{36} & \frac{11}{36} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

1) Test of Significance for Single mean:

- $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

- Confidence Interval for mean is

$$\left(\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

2) Test of Significance for difference of two means:

- $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ $\quad (\sigma_1^2 \neq \sigma_2^2)$

$= \dfrac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ $\quad (\sigma_1^2 = \sigma_2^2 = \sigma)$

3) Test of Significance for Single Proportion:

- $Z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}}$

p = Sample Proportion
$= \frac{x}{n}$
P = Population Proportion
$Q = 1 - P$

- Confidence Interval for Single Proportion

$$\left(p - Z_{\alpha/2}\sqrt{\frac{PQ}{n}}, p + Z_{\alpha/2}\sqrt{\frac{PQ}{n}}\right)$$

If proportion is not given,

$$\left(p - Z_{\alpha/2}\sqrt{\frac{pq}{n}}, p + Z_{\alpha/2}\sqrt{\frac{pq}{n}}\right)$$

4) Test of Significance for difference of two proportions:

$$Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad P = \frac{n_1P_1 + n_2P_2}{n_1 + n_2} \qquad Q = 1 - P$$

* If only population proportions are known but sample proportions are not known then test statistic is

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}}$$

* Confidence Interval is $\left(P_1 - Z_{\alpha/2}\sqrt{\frac{P_1Q_1}{n_1}}, P_1 + Z_{\alpha/2}\sqrt{\frac{P_1Q_1}{n_1}}\right)$

UNIT-IV

t-test : (E-test)

- The test statistic for t-test (single mean) is

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} \qquad \text{[If S.D is known]}$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \qquad \text{[If S.D is unknown]}$$

$$\boxed{\nu = n-1}$$

- For Two-tailed Test $LOS = \frac{\alpha}{2}$

  One-tailed Test $LOS = \alpha$

* Confidence interval for single mean is

$$\left(\bar{x} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right), \bar{x} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)\right)$$

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

- The test statistic for difference of two means (t-test) is

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{s^2\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}}$$

where $s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}$ (If S.D is known)

$$s^2 = \frac{1}{n_1+n_2-2}\left[\sum(x_i-\bar{x})^2 + \sum(y_i-\bar{y})^2\right] \text{ (If S.D is unknown)}$$

$$\boxed{V = n_1+n_2-2}$$

F-test:

$$F = \frac{\text{Greater Variance}}{\text{Smaller Variance}}$$

$$F = \frac{S_1^2}{S_2^2} \quad (\text{if } S_1^2 > S_2^2)$$

$$F = \frac{S_2^2}{S_1^2} \quad (\text{if } S_1^2 < S_2^2)$$

$$\left.\begin{aligned} S_1^2 &= \frac{1}{n_1-1}\sum(x_i-\bar{x})^2 \\ S_2^2 &= \frac{1}{n_2-1}\sum(y_i-\bar{y})^2 \end{aligned}\right\} \text{If sum of squares of deviation are known}$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1}$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1}$$

If only S.D of sample is known

$$\boxed{\begin{array}{l} V_1 = n_1 - 1 \\ V_2 = n_2 - 1 \end{array}}$$

* If $S_1^2 > S_2^2 \Rightarrow (V_1, V_2)$
* If $S_2^2 > S_1^2 \Rightarrow (V_2, V_1)$

corresponding values should be considered

* Come from the same normal Population $\Rightarrow$

When we accept null hypothesis in both F-test and T-test

* <u>Chi-Square (χ^2-test):</u>

* $$\chi^2 = \frac{(O_i - E_i)^2}{E_i}$$

$O_i \rightarrow$ Observed Frequency

$E_i \rightarrow$ Expected Frequency

$$E_i = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

$$V = (\text{no. of rows} - 1)(\text{no. of columns} - 1)$$

Karl Pearson Correlation Coefficient:

- Also known as Product – Moment correlation coefficient
- $$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$ $x = x - \bar{x}$, $y = y - \bar{y}$
- Correlation coefficient limits are : $-1 \leq r \leq 1$
- If $r = 1$, correlation is perfect and positive
- If $r = -1$, correlation is perfect and negative
- If $r = 0$, there is no relationship between the two variables

Spearman's Rank Correlation Coefficient:

* When the ranks are given (not repeated)

$$\rho = 1 - \frac{6 \sum D^2}{N(N^2-1)}$$

D = First Rank – Second Rank $= R_1 - R_2$

* When the ranks are not given (repeated)

$$\rho = 1 - 6\left\{\frac{\sum D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \cdots}{N(N^2-1)}\right\}$$

where $m_1, m_2 \ldots$ are ranks which are repeated

4) **Regression:** It is a statistical method which helps us to find the values of unknown variable using the value of known variable.

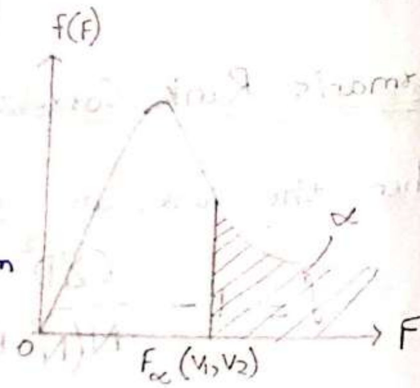
* The equation of line of regression of X on Y is

$$x - \bar{x} = b_{xy}(y - \bar{y}) \qquad b_{xy} = \frac{\sum xy}{\sum y^2}$$

* The equation of line of regression of Y on X is

$$y - \bar{y} = b_{yx}(x - \bar{x}) \qquad b_{yx} = \frac{\sum xy}{\sum x^2}$$

Properties of f-distribution:



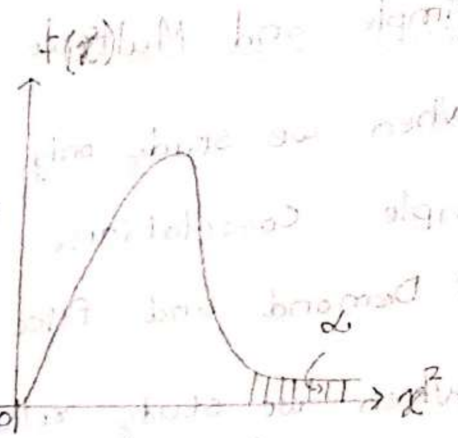
1) F-distribution is free from population parameters and depends upon degrees of freedom only.

2) F-distribution curve lies entirely in first quadrant.

3) F-distribution depends not only on the two parameters ν_1 and ν_2 but also on the order in which they are stated.

4) The mode of f-distribution is less than unity.

Properties of χ^2 Distribution:



1) χ^2 - distribution curve is not symmetrical, lies entirely in the first quadrant.

2) It depends only on the degrees of freedom V

3) If χ_1^2 and χ_2^2 are two independent distributions with v_1 & v_2 degrees of freedom, $\chi_1^2+\chi_2^2$ will be χ^2 - squared distribution with (V_1+V_2) degrees of freedom i.e it is additive.

Types of Correlation:

① Positive and Negative Correlation:

* It depends upon the direction of change of the variables.

* If two variables move together in the same direction an increase in the value of one variable is accompanied by an increase in the value of the other variable (or) decrease in one variable is accompanied by the decrease in another variable, then the correlation is Positive.

Eg: Height & Weight, Price & Supply

* If two variables move together in opposite directions, then the correlation is negative.

② Simple and Multiple Correlation:

* When we study only two variables, the relationship is simple Correlation.

Eg: Demand and Price

* When we study relationship between more than two variables, it is Multiple Correlation.

Eg: Price, Demand and Supply

③ Partial and Total Correlation:

* The study of two variables excluding some other variables is called Partial Correlation.

* In total correlation, all the facts are taken into accounts

④ Linear and Non-Linear Correlation:

* If the ratio of change between two variables is uniform then there will be a linear correlation between them.

* In Non-linear correlation the amount of change in one Variable doesnot bear a constant ratio to the amount of change in the other variables.

$C.C = \sqrt{b_{xy} \times b_{yx}}$

* $\theta = \left(\frac{1-r^2}{r}\right)\left[\frac{\sigma_x \sigma_y}{\sigma_x^2+\sigma_y^2}\right]$

Discrete Random Variable:

$$\mu = \sum_{i=1}^{n} x_i P_i$$

$$\sigma^2 = \sum_{i=1}^{n} x_i^2 P_i - \mu^2$$

Continuous Random Variable:

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

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

Binomial Distribution	Poisson Distribution
$P(x) = {}^nC_x\, p^x q^{n-x}$	$P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$
$\mu = np$	$\mu = \lambda$
$\sigma^2 = npq$	$\sigma^2 = \lambda$

Normal Distribution:

* $Z = \dfrac{x-\mu}{\sigma}$

* If Z_1 & Z_2 are of same sign then

$$P(Z_1 \le Z \le Z_2) = |A(Z_2) - A(Z_1)|$$

* If Z_1 & Z_2 are of opposite sign then

$$P(Z_1 \le Z \le Z_2) = |A(Z_2) + A(Z_1)|$$

~~Bino~~Normal Approximation to Binomial Distribution:

$$Z_1 = \frac{(x_1 - \frac{1}{2}) - \mu}{\sigma} \qquad Z_2 = \frac{(x_2 + \frac{1}{2}) - \mu}{\sigma}$$