UNIT-1 # Probability

→ If an experiment is repeated under similar and identical conditions, we generally come across 2 types of situations.

1. The net result, what is generally known as outcome is unique (or) certain, these situations are known as deterministic (or) predictable.
2. The net result, is not unique but may be 1 on of several possible outcomes, these are known as probabilistic (or) unpredictable.

(1) Random Experiment: If an exp is conducted any no. of times under essentially identical conditions, there is a set of all possible outcomes associated with it. If the result is not certain and is any of the several possible outcomes, is known as Random Experiment (or) a Random trial.

→ The outcomes in the random exp are known as Events.

→ An event in a trial that cannot be further split is called a simple event (or) an elementary event.

→ The set of all possible simple events in a trial is called Sample space for the trial. and is denoted by 'S'

Equall Likely Events:

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

Ex: when a card is drawn from the pack any ~~pac~~ card may be obtained, in this trial all 52 elementary events are equally likely

Exhaustive Events: All possible events in any trial are known as exhaustive events.

Ex In throwing a die there are 6 Exhaustive events, they are: getting 1 or 2 or 3 or 4 or 5 or 6. are the Exhaustive events.

* Mutually Exclusive Events: The events are said to be mutually exclusive, if no 2 (or) more ~~are~~ of the events can happen simultaneously in the same trial.

* Definition of Probability:

In a random experiment, let there be n mutua exclusive and equally likely events, let 'E' be the event of the experiment. If 'm' elementary events which are favarable to 'E' then the probability of happening of events 'E' is given by ~~PE~~

$$P(E) = \frac{m}{n} = \frac{\text{no. of favarable events in } E}{\text{total no. of possible events in } S}$$

→ If $\bar{E}$ denotes the event of non-occurence of E, the no. of events in $\bar{E}$ is $\boxed{n-m}$

$$\therefore P(\bar{E}) = \frac{n-m}{n}$$

$$P(\bar{E}) = 1 - \frac{m}{n}$$

$$P(\bar{E}) = 1 - P(E)$$

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

also, we have $0 \leq P(E) \leq 1$ and

$$0 \leq P(\bar{E}) \leq 1$$

Axioms of probability:

1. Axiom of positivity: $P(E) \geq 0$ for all the events of E of S
2. Axiom of certainity: $P(S) = 1$
3. Axiom of Union: If E_1 & E_2 are the events of S then

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

If $E_1, E_2, E_3 \ldots, E_n$ are the events of S then

$$P(E_1 \cup E_2 \cup E_3 \cup \ldots \cup E_n) = P(E_1) + P(E_2) \ldots P(E_n)$$

* Complementary Events:

Two events of a sample space (S) whose Union(U) is entire Sample space whose intersection is ϕ then, the two events are called as Complementary events to each other.

i.e, $E \cup \bar{E} = S$

$E \cap \bar{E} = \phi$ then, E & $\bar{E}$ are said to be Complementary events.

Compound Event: when 2 (or) more events occur in conjunction with each other, their joint occurence is called compound event.

* Addition theorem on Probability:

If E_1 & E_2 are 2 events in Sample space (S) then

$$P(E_1 \text{ or } E_2) = P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Proof

Let E_1, E_2 are the two events in sample space 'S'

Case-1: when $E_1 \cap E_2 \neq \emptyset$

$$E_1 = \{a_1 a_2 \ldots a_k, a_{k+1} \ldots a_{k+l}\}$$

$$E_2 = \{a_{k+1} a_{k+2} \ldots a_{k+l} a_{k+l+1} \ldots a_{k+l+m}\}$$

$$E_1 \cap E_2 = \{a_{k+1} a_{k+2} \ldots a_{k+l}\}$$

$$E_1 \cup E_2 = \{a_1 a_2 \ldots a_k a_{k+1} \ldots a_{k+l} a_{k+l+1} \ldots a_{k+l+m}\}$$

$$\Rightarrow P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

$$= P\{a_1 a_2 \ldots a_k a_{k+1} \ldots a_{k+l}\} + P\{a_{k+1} a_{k+2} \ldots a_{k+l} a_{k+l+1} \ldots a_{k+l+m}\}$$

$$- P\{a_{k+1} a_{k+2} \ldots a_{k+l}\}$$

$$= \{P(a_1) + P(a_2) + \ldots P(a_k) + P(a_{k+1}) \ldots P(a_{k+l}) + P(a_{k+1}) + P(a_{k+2})$$

$$\ldots P(a_{k+l}) + P(a_{k+l+1}) \ldots P(a_{k+l+m})$$

$$- P(a_{k+1}) - P(a_{k+2}) \ldots - P(a_{k+l+m})\}$$

$= P\{a_1 a_2 - a_k\, a_{k+l} \cdots a_{k+l+m}\}$

$= P(E_1 \cup E_2)$

$\therefore P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

Case-2: when $E_1 \cap E_2 = \emptyset$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - 0$$

$$= P(E_1) + P(E_2) - P(\emptyset)$$

$$\therefore P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

→ If E_1, E_2 & E_3 are the 3 events in sample space S then

$$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_2 \cap E_3)$$
$$- P(E_1 \cap E_3) + P(E_1 \cap E_2 \cap E_3)$$

① A card is drawn from well shuffled pack of cards then what is the probability that is either spade (or) an Ace

Sol Let S be the sample space.

Let A denotes the event of getting a spade $P(A) = \frac{13C_1}{52C_1}$

Let B be the event of getting an ACE is $P(B) = \frac{4C_1}{52C_1}$

Let $A \cap B$ be the event of getting a spade and an ace is

$P(A \cap B) = \frac{1C_1}{52C_1}$

Probability of getting a Ace (or) a spade.

$P(\text{spade or Ace}) = P(A \text{ or } B)$

$= P(A \cup B)$

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

$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52}$

$= \frac{16}{52} = 0.3077$

(2) Three students A, B, C are in a running race A & B have the same probability of winning & each is twice as likely to win as C. then find the probability that B & C win

Sol A, B, C are in a sample space.

$P(A) = P(B) = 2P(C)$

Since A, B, C are in sample space.

$A \cup B \cup C = S$

$P(A \cup B \cup C) = P(S)$

$P(A) + P(B) + P(C) = P(S)$

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

$5P(C) = 1$

$P(C) = 1/5$

$\Rightarrow P(A) = 2/5$

$P(B) = 2/5$

$P(B \text{ or } C) = P(B \cup C)$

$= P(B) + P(C)$

$= 2/5 + \frac{1}{5}$

$\therefore P(B \cup C) = 3/5 = 0.6.$

(3) From a city 3 Newspaper A, B, C are being published. A is read by 20%, B is read by 16% & C is read by 14% both A & B are readed by 8% both B and C are red by 5% Both C & A are read by 40% and all the 3 A, B C are read by 2% then the % of the population that read about atleast one paper

Sol ~~P(A~~

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

$P(B) = \cancel{20}16\% = 0.16$

$P(C) = \cancel{0}14\% = 0.14$

$P(A \cap B) = 8\% = 0.08$

$P(B \cap C) = 5\% = 0.05$

$P(A \cap C) = 4\% = 0.04$

$P(A \cap B \cap C) = 2\% = 0.02$

By addition theorem on three events.

$P(A \text{ or } B \text{ or } C) = P(A \cup B \cup C)$

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

$= 0.2 + 0.16 + 0.14 - 0.05 - 0.04 - 0.8 + 0.02$

$= 0.35$

$\therefore$ The pecentage of the population that read atleast one newspaper $\Rightarrow P(A \cup B \cup C) \times 100\%$

$= (0.35)(100) = \underline{35\%}$

* Conditional Event:

If E_1, E_2 are the event of the sample space S and If E_2 occurs after the occurance of E_1 then the event of occurance of E_2 after E_1 is called Conditional event of E_2 given by E_1 and is denoted by $\frac{E_2}{E_1}$

Similarly, we define the conditional event of E_1 after E_2 given by $\frac{E_1}{E_2}$

* Conditional Probability:

If E_1 and E_2 are two event of a sample space 'S' and ~~P(E~~ $P(E_1 \neq 0)$ then the probability of E_2 after E_1 has occured is called Conditional probability of E_2 given by E_1

→ It is defined as $P\left(\frac{E_2}{E_1}\right) = \frac{P(E_1 \cap E_2)}{P(E_1)}$ where $P(E_1) \neq 0$

we also defined conditional probability of E_1 given by E_2 i

$$P\left(\frac{E_1}{E_2}\right) = \frac{P(E_1 \cap E_2)}{P(E_2)} \text{ where } P(E_2) \neq 0$$

Eg when 2 coins are tossed, the event of getting two heads. Given that their is atleast one head is a conditional event

Dependent & Independent Events:

→ If the occurance of event E_2 is not affected by the accurance of E_1 then the event of E_2 is said to independent of E_1 and $P\left(\frac{E_2}{E_1}\right) = P(E_2)$

Pairwise independent Events:

Let E_1, E_2, E_3 are the event of sample space 'S' they are said to be pairwise independent If

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

$P(E_2 \cap E_3) = P(E_2) \cdot P(E_3)$

$P(E_1 \cap E_3) = P(E_1) \cdot P(E_3)$

where $P(E_1) \neq 0$

$P(E_2) \neq 0$

$P(E_3) \neq 0$

* Multiplication Theorem of Probability:

In a random experiment if E_1, E_2 are the two event such that $P(E_1) \neq 0, P(E_2) \neq 0$ then

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

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

Let 'S' be the sample space associated with a random experiment.

Let E_1, E_2 by the event of S such that $P(E_1) \neq 0$ & $P(E_2) \neq 0$

Since. $P(E_1) \neq 0$ ~~by the event of S such that~~

by the definition conditional probability of E_2 given by E_1

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

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

Since, $P(E_2) \neq 0$ by the definition of conditional probability of E_1 given by E_2

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

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

① Determine (i) $P(B/A)$ · (ii) $P\left(\frac{A}{B^c}\right)$, (iii) $P\left(\frac{B}{A^c}\right)$ if A & B are the events $\Rightarrow P(A) = \frac{1}{3}$, $P(B) = \frac{1}{4}$, $P(A \cup B) = 1/2$

Sol Given that.

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

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

$$= \frac{1}{3} + \frac{1}{4} - \frac{1}{2}$$

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

(i) $P(B/A) = \frac{P(A\cap B)}{P(A)} = \frac{1/12}{1/3} = \frac{1}{4}$

(ii) $P\left(\frac{A}{B^c}\right) = \frac{P(A\cap B^c)}{P(B^c)}$ $\Rightarrow P(B^c) = 1 - P(B)$

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

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

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

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

$= \frac{1}{4}$

$P\left(\frac{A}{B^c}\right) = \frac{1/4}{3/4} = \frac{1}{3}$

(iii) $P\left(\frac{B}{A^c}\right) = \frac{P(B\cap A^c)}{P(A^c)}$ $\Rightarrow P(A^c) = 1 - P(A)$

$= 1 - 1/3 = \underline{\underline{2/3}}$

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

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

$P(B\cap A^c) = 1/6$

$\therefore P\left(\frac{B}{A^c}\right) = \frac{1/6}{2/3} = 1/4$

② A can hit the target 3 in 5 times, B hits the target 2 in 5 times and C hits the target 3 in 4 times then find the probability of the target being hits when all of the try.

Sol Given that,

Let $P(A) = 3/5$ $\quad P(B) = 2/5$ $\quad P(C) = 3/4$

$P(A^c) = 1 - P(A)$ $\quad \Rightarrow P(B^c) = 1 - P(B)$ $\quad \Rightarrow P(C^c) = 1 - P(C)$

$= 1 - 3/5$ $\quad = 1 - \frac{2}{5}$ $\quad = 1 - 3/4$

$P(A^c) = 2/5$ $\quad \therefore P(B^c) = 3/5$ $\quad \therefore P(C^c) = 1/4$

the probability that name of A, B, C hits the target

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

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

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

$= 3/50 = 0.06$

$\therefore$ The probability of atleast one of A, B, C hitting the target, $P(A \cup B \cup C)$

$\therefore P(A \cup B \cup C) = 1 - \left[P(A^c)\, P(B^c)\, P(C^c)\right]$

$= 1 - 0.06$

$\therefore P(A \cup B \cup C) = 0.94$

* 3 machines A, B, C produce 40%, 30% & 30% res of the total no. of items of the factory. The percentages of defective items of the machines are 4%, 2% & 3% respectively. In an item is selectively at random then find the probabi that the item is defective

Sol Given that

$$P(A) = 40\% = 0.4$$
$$P(B) = 30\% = 0.3$$
$$P(C) = 30\% = 0.3$$

probability of defective item from machine A $P\left(\frac{D}{A}\right) = 4\% = 0.04$

probability of defective item from machine B, $P\left(\frac{D}{B}\right) = 2\% = 0.02$

probability of defective item from machine C, $P\left(\frac{D}{C}\right) = 3\% = 0.03$

– ∴ probability of defective item,

$$P(\text{defective}) = P(D) =$$

$$= P(D) = P\{(D\cap A) \text{ or } (D\cap B) \text{ or } (D\cap C)\}$$

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

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

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

$$P(D) = \underline{0.031}$$

∴ Probability of defective item from A, B, C is 0.031.

* **Baye's Theorem:**

Statement:

Let $E_1, E_2, \ldots, E_n$ are 'n' mutually exclusive event and exhaustive events such that $p(E_i > 0)$ $\forall$ the events in a sample spase 'S' and A is any other event in S intersecting with every E_i (i.e, A can occur in combination with any one of the events $E_1, E_2, \ldots E_n$) such that $P(A) > 0$

In E_i is any of the events of $E_1, E_2, \ldots E_n$ where $P(E_1), P(E_2) \ldots P(E_n)$ and $P\left(\frac{A}{E_1}\right), P\left(\frac{A}{E_2}\right) \ldots P\left(\frac{A}{E_n}\right)$ are known then,

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

* Let $E_1, E_2, \ldots E_n$ are 'n' mutually exclusive and exhaustive events of 'S' such that $P(E_i) > 0$

A is any other event of S where $P(A) > 0$

$S = \{E_1, E_2, \ldots E_n\}$

i.e, $S = \{E_1 \cup E_2 \cup \ldots \cup E_n\}$

$A = A \cap E_i$

$A = A \cap S$

$A = A \cap \{E_1 \cup E_2 \cup \ldots \cup E_n\}$

~~P(A) = 1~~

$A = (A \cap E_1) \cup (A \cap E_2) \cup \ldots \cup (A \cap E_n)$

$P(A) = P\{(A \cap E_1) \cup (A \cap E_2) \cup \ldots \cup (A \cap E_n)\}$

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \cdots + P(A \cap E_n)$$

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

here, $P(E_1), P(E_2), \ldots, P(E_n)$ and

$P\left(\frac{A}{E_1}\right), P\left(\frac{A}{E_2}\right), \ldots, P\left(\frac{A}{E_n}\right)$ are known probabilities

→ By the definition of conditional probability of the Event (E_k) given by 'A' is,

$$P\left(\frac{E_K}{A}\right) = \frac{P(A \cap E_K)}{P(A)} \quad \text{where } P(A) \neq 0$$

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

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

$$P\left(\frac{E_K}{A}\right) = \frac{P(E_K) \cdot P\left(\frac{A}{E_K}\right)}{\sum_{i=1}^{n} P(E_i) \cdot P\left(\frac{A}{E_i}\right)}$$

NOTE:

* Baye's theorem is also known as Formulae for the Probability of Causes i.e, Probability of a particular cause E_i, given that the event 'A' has happened already

* here, $P(E_i)$ is a priority probability known even before the experiment,

$P(\frac{A}{E_i})$ called likelyhoods and $P(\frac{E_i}{A})$ is called Posteriori probability which determined after the result of the experiment.

Problems on Baye's theorem:

In a certain college 25% of boys and 10% of girls are studying mathematics. the girls constitute 40% of the student body.

(i) What is the probability that mathematics is being studie

(ii) If a student is selected at random, and is found to be studying mathematics, find the probability that the student is a boy.

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

Sol Probability of boys, $P(B) = 60\% = 0.6$

Probability of girls, $P(G) = 40\% = 0.4$

probability of mathematics student being a boy, $P(\frac{M}{B}) = 25\% = 0.25$

probability of mathematics student being a ~~boy~~ Girl, $P(\frac{M}{G}) = 10\% = 0.1$

(i) Probability of mathematics is being studied, $P(M)$

$$P(M) = P\{(M \cap B) \text{ or } (M \cap G)\}$$

$$= P(M \cap B) \cup P(M \cap G)$$

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

$$P(M) = (0.6)(0.25) + (0.4)(0.1)$$

$$\therefore P(M) = \underline{0.19}$$

(ii) If a student is selected at random and is found to be studying mathematics,

probability of the student is a boy. $P\left(\frac{B}{M}\right)$

$$P\left(\frac{B}{M}\right) = \frac{P(B \cap M)}{P(M)}$$

$$= \frac{P(B) \cdot P\left(\frac{M}{B}\right)}{P(M)}$$

$$= \frac{(0.6)(0.25)}{0.19}$$

$$\therefore P\left(\frac{B}{M}\right) = \underline{0.7895}$$

(iii) If a student is selected at random and is found to be studying mathematics,

probability of the student is a girl, $P\left(\frac{G}{M}\right)$

$$P\left(\frac{G}{M}\right) = \frac{P(G \cap M)}{P(M)}$$

$$= \frac{P(G) \cdot P\left(\frac{M}{G}\right)}{P(M)}$$

$$= \frac{(0.4) \cdot (0.1)}{0.19}$$

$$P\left(\frac{G}{M}\right) = \underline{0.2105}$$

② In a bolt factory Machines, A, B, C manufactures 20%, 30% & 50% ~~res~~ of the total of their output & 6%, 3%, 2% are defective.

A bolt is drawn at random and is found to be defective, find the probability that it is manufactured

from
(i) Machine A ② Machine B ③ Machine C

Sol Given that

probability of A $P(A) = 20\% = 0.2$

probability of B, $P(B) = 30\% = 0.3$

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

probability of defective bolt from Machine A, $P\left(\frac{D}{A}\right) = 6\% = 0.06$

probability of defective bolt from Machine B, $P\left(\frac{D}{B}\right) = 3\% = 0.03$

" " " " " Machine C $P\left(\frac{D}{C}\right) = 2\% = 0.02$

$\Rightarrow$ Probability of defective bolt $= P(D)$

$$P(D) = P\{(D \cap A) \text{ or } (D \cap B) \text{ or } (D \cap C)\}$$

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

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

$$= (0.2)(0.06) + (0.3)(0.03) + (0.5)(0.02)$$

$$P(D) = 0.031$$

(i) probability of defective bolt that is manufactured from Machine A

$$P\left(\frac{A}{D}\right) = \cancel{P\{(A \cap D) \text{ or }} (B.$$

$$= \frac{P(A \cap D)}{P(D)}$$

$$= \frac{P(A) \cdot P\left(\frac{D}{A}\right)}{P(D)}$$

$$= \frac{(0.2)(0.06)}{0.031}$$

$$P\left(\frac{A}{D}\right) = 0.3871$$

(ii) from machine B

$$P\left(\frac{B}{D}\right) = \frac{P(B\cap D)}{P(D)}$$

$$= \frac{P(B)\cdot P\left(\frac{D}{B}\right)}{P(D)}$$

$$= \frac{(0.3)(0.03)}{0.031}$$

$$P\left(\frac{B}{D}\right) = 0.2903$$

(iii) from machine C

$$P\left(\frac{C}{D}\right) = \frac{P(C\cap D)}{P(D)}$$

$$= \frac{P(C)\cdot P\left(\frac{C}{D}\right)}{P(D)}$$

$$= \frac{(0.5)(0.02)}{0.031}$$

$$\therefore P\left(\frac{C}{D}\right) = 0.3226$$

(3) Suppose 5 men out of 100 and 25 women out of 10,000 are colour blind, A colour blind person is chosen at random then, what is the probability that the colour blind person being a male, also, obtain the probability of ~~col~~ female having the colour blind. (Assume that probabilites of male or female are equal).

Sol Given that;

Probability of colourblind person being a Male, $P\left(\frac{C}{M}\right) = \frac{5}{100} = 0.05$

Probability of colourblind person being a female, $P\left(\frac{C}{F}\right) = \frac{25}{10000} = 0.0025.$

Probability of ~~men~~ male and female $= P(M) = P(F) = \frac{1}{2}$

$$\therefore P(M) = P(F) = 0.5$$

Probability of colour blind person $P(C)$

$$P(C) = P\{(C \cap M) \text{ or } (C \cap F)\}$$

$$= P(C \cap M) \cup P(C \cap F)$$

$$= \cancel{P(C) \cdot P\left(\frac{M}{C}\right)}$$

$$P(M) \cdot P\left(\frac{C}{M}\right) + P(F) \cdot P\left(\frac{C}{F}\right)$$

$$= 0.5 \times 0.05 + (0.0025)(0.5)$$

$$P(C) = 0.0262$$

(i) Probability of male is a colour blind person, $P\left(\frac{M}{C}\right)$

$$P\left(\frac{M}{C}\right) = \frac{P(M \cap C)}{P(C)}$$

$$= \frac{P(M) \cdot P\left(\frac{C}{M}\right)}{P(C)}$$

$$= \frac{(0.5)(0.05)}{0.0262}$$

$$P\left(\frac{M}{C}\right) = 0.9542$$

(ii) Probability of female is a colour blind person, $P\left(\frac{F}{C}\right)$

$$P\left(\frac{F}{C}\right) = \frac{P(F \cap C)}{P(C)}$$

$$= \frac{P(F) \cdot P\left(\frac{F}{C}\right)}{P(C)}$$

$$= \frac{(0.5)(0.0025)}{0.0262}$$

$$P\left(\frac{F}{C}\right) = 0.0477$$

(4) Of the 3 men, the chances that a politician, a business or an acadamician will be appointed as a 'VC' of a University are 0.5, 0.3, 0.2 respectively. Probability that reasearch is promoted by these persons, if they are appointed as VC are 0.3, 0.7, 0.8 respectively.

(i) Determine the probability that the research is promoted

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

Sol Given that: Let A, B, C are the events of politician, business, Acad...

probability of politician, $P(A) = 0.5$

probability of business man, $P(B) = 0.3$

Probability of acamedician, $P(C) = 0.2$

→ Probability of research promoted by politician. $P\left(\frac{R}{A}\right) = 0.3$

prd by businessman, $P\left(\frac{R}{B}\right) = 0.7$

by acadamecian, $P\left(\frac{R}{C}\right) = 0.8$

(i) Probability of research promoted

$$P(R) = P\{(R \cap A) \text{ or } (R \cap B) \text{ or } (R \cap C)\}$$

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

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

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

$$P(R) = 0.52$$

If research is promoted

(ii) Probability of VC is an academician $P\left(\frac{C}{R}\right)$

$$P\left(\frac{C}{R}\right) = \frac{P(C \cap R)}{P(R)}$$

$$= \frac{P(C) \cdot P\left(\frac{R}{C}\right)}{P(R)}$$

$$= \frac{(0.8)(0.2)}{0.52}$$

$$P\left(\frac{C}{R}\right) = 0.3077$$

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

(i) Hotel X (ii) Hotel Y , (iii) Hotel Z.

Sol Given

probability of hotel X, $P(X) = 20\% = 0.2$

" hotel Y, $P(Y) = 50\% = 0.5$

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

probability of faulty plumbing roon from hotel X, $P\left(\frac{F}{X}\right) = 5\%$
$= 0.05$

from hotel Y, $P\left(\frac{F}{Y}\right) = 4\%$
$= 0.04$

$P\left(\frac{F}{Z}\right) = 8\% = 0.08$

probability of faulty plumbing rooms:

$$P(F) = P\{(F \cap X) \text{ or } (F \cap Y) \text{ or } (F \cap Z)\}$$

$$= P(F \cap X) \cup P(F \cap Y) \cup P(F \cap Z)$$

$$= P(X) \cdot P\left(\frac{F}{X}\right) + P(Y) \cdot P\left(\frac{F}{Y}\right) + P(Z) \cdot P\left(\frac{F}{Z}\right)$$

$= (0.2)(0.05) + (0.5)(0.04) + (0.3)(0.08)$

$P(F) = 0.054$

(i) Probability that the businessman's room having faulty plumbing is assigned to hotel X:

$$P\left(\frac{X}{F}\right) = \frac{P(X\cap F)}{P(F)}$$

$$= \frac{P(X)\cdot P\left(\frac{F}{X}\right)}{P(F)}$$

$$= \frac{(0.2)(0.05)}{0.054}$$

$$P\left(\frac{X}{F}\right) = 0.1859$$

(ii) hotel Y:

$$P\left(\frac{Y}{F}\right) = \frac{P(Y\cap F)}{P(F)}$$

$$= \frac{P(Y)\cdot P\left(\frac{F}{Y}\right)}{P(F)} = \frac{(0.5)(0.04)}{0.054} = 0.3703$$

$$\therefore P\left(\frac{Y}{F}\right) = 0.3703$$

(iii) hotel Z:

$$P\left(\frac{Z}{F}\right) = \frac{P(Z\cap F)}{P(F)}$$

$$= \frac{P(Z)\cdot P\left(\frac{F}{Z}\right)}{P(F)} = \frac{(0.3)(0.08)}{0.054}$$

$$P\left(\frac{Z}{F}\right) = 0.4444$$

6) A bag A contains 2 white and 3 red balls & in bag B contains 4 white and 5 red balls; 1 ball is drawn at random from one of the bags and it is found to be red, find the probability that the red ball is drawn (i) from bag B;
(ii) & from bag A

Sol Given that:

probability of red ball from bag A $P\left(\frac{R}{A}\right) = \frac{3}{5} = 0.6$

probability of red ball from bag B $P\left(\frac{R}{B}\right) = \frac{5}{9} = 0.5555$

NOTE: Probability of bag A & P(A) = probability of P(B)

probability of red ball & P(R) =

$$P\{(R \cap A) \text{ or } (R \cap B)\}$$

$$= P(R \cap A) \cup P(R \cap B)$$

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

$$= (0.5) \cdot (0.6) + (0.5) \cdot (0.555)$$

$$P(R) = \cancel{0.575}$$
$$= 0.5775$$

(i) Probability of red ball from bag B

$$P\left(\frac{B}{R}\right) = \frac{P(B \cap R)}{P(R)}$$

$$= \frac{P(B) \cdot P\left(\frac{R}{B}\right)}{P(R)}$$

$$= \frac{(0.5) \cdot (0.5555)}{0.5775}$$

$$P\left(\frac{B}{R}\right) = 0.4808$$

(ii) from bag A

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

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

$$= \frac{(0.5) \cdot (0.6)}{0.57775}$$

$$\underline{P\left(\frac{A}{R}\right) \simeq 0.5192}$$

* Random Variables and distribution functions:

* Real variable 'x' whose value is determined by the outcome of the random exp is called a Random variable.

→ the random variables are of 2 types.

1. Discrete Random variables.
2. Continuous Random variable.

(1) Discrete Random variable:

A random variable 'x' which can take only a finite no. of discrete values in an interval of domain, is called a Discrete Random variable

(or)

The random variables takes the values only on the set $\{0, 1, 2, 3, \ldots\ldots n\}$ is called A discrete Random variable

E.g: No. of telephone calls recieved by a telephone operator

No. of printmistakes in each of a textbook.

2. Continuous Random Variable:

A random variable 'x' which can takes the values continuously

E.g: Age of individual, temperature, & time, etc.

Probability distribution functions:

* Let 'x' be a random variable, the probability distribu function associated with 'x' is defined as the probability that the outcome of an experiment will be one of the outcomes.

→ The function $F_X(x)$ is defined by $F_X(x) = P(X \leq x)$

→ If the no.'s $p(x_i)$ for $i = 1, 2, 3 \ldots$ satisfies the 2 condition $P(x_i) \geq 0 \; \forall$ the values of i

$\sum_{i=1}^{n} P_i = 1$ then the function P is called probability mass function of the random variable)

* Mean (or) Expectation, variants and standard deviation of a probability distribution.

1. Expectation:

Suppose a random variable 'x' assumes the values $x_1, x_2 \ldots x_n$ w.respective probabilities $P_1, P_2, \ldots P_n$ then the Expectation (or) Expected value (or) mean of X is denoted $E(X)$ and is defined as sum of product of diff values of x and their corresponding probabilities

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

in general, the expected value of a func $g(x)$ is defined by $\boxed{E[g(x)] = \sum_{i=1}^{n} P_i \cdot g(x_i)}$

② Variants:

Variance:

Variance characterizes the variability in the distribution since 2 distributions with same mean can still have different dispersions of data about their means.

→ The variance (σ^2) of the discrete distribution function is given by

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

$$= \sum (x_i^2+\mu^2-2x_i\mu) \cdot P_i$$

$$= \sum x_i^2 P_i + \sum \mu^2 \cdot P_i - 2\sum x_i \mu P_i$$

$$= E(X^2) + \mu^2 \sum P_i - 2\mu \sum x_i P_i$$

$$= E(x^2) + \mu^2(1) - 2\mu \cdot \mu$$

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

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

* Standard deviation:

Standard deviation is the +ve sqrt of the variance

i.e, $\sigma = \sqrt{E(x^2) - [E(x)]^2}$

(or)

$\sigma = \sqrt{\sum x_i^2 \cdot P_i - \mu^2}$

* Results on Expectation:

① If X is a random variable and k is a constant then $E(X+k) = E(x)+k$

Proof: By the definition of expectation

we have $E(x) = \sum x_i p_i$

$$E[x+k] = \sum(x_i+k)p_i$$

$$= \sum x_i p_i + \sum k p_i$$

$$= E[x] + k\sum p_i$$

$$= E(x) + k(1)$$

$$\Rightarrow E[x+k] = E(x)+k$$

② If X & Y are the two discrete random variables then $E(X+Y) = E(x)+E(Y)$ provided that $E(X)$ and $E(Y)$ are exists.

Proof By the definition of expectation,

we have, $E(x) = \sum_{i=1}^{n} p_i x_i$

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

Let p_{ij} be the joint probability defined as,

$$p_{ij} = P(X=x_i)\,\&\,(Y=y_i)$$

$$= P(X=x_i)\cap(Y=y_j)$$

$$= P(x_i)\cap P(y_j)$$

$$= p_i \cap p_j$$

$$= p_i \cdot p_j$$

$$= p_i p_j$$

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

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

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} \left[P_i \cdot P_j \, x_i + P_i \cdot P_j \cdot y_j\right]$$

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

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

$$\boxed{E(X+Y) = E(X) + E(Y)}$$

③ If 'x' & 'y' are 2 discrete and independent random variables then

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

Proof: By the definition

$$E(X) = \sum_{i=1}^{n} P_i X_i \qquad E(Y) = \sum_{i=1}^{m} P_i y_i$$

$$P_{ij} = P(X = x_i \ \& \ Y = y_i)$$

$$= P(x_i) \cap P(y_i)$$

$$= \cancel{P(x_i)} \; P_i \cap P_j$$

$$= P_{ij}$$

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

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} (x_i y_i) \cdot P_i \cdot P_j$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \left[P_i P_j x_i \cdot P_i P_j y_i \right]$$

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

$$= E(X)(1) \cdot E(Y)(1)$$

$$E(x \cdot y) = E(x) \cdot E(y)$$

① If X is a random variable and who's probability function is given by.

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

Determine:

(i) The value of k

(ii) $P(1 \le x < 7)$

(iii) if $P(x \le k) > \frac{1}{2}$ then, find the minimum value of k

(iv) probability distribution func of 'x'

(v) Mean (vi) Variance (vii) Standard deviation.

sol (i) We know that

$$\sum_{i=1}^{7} P_i = 1$$

$$P_0 + P_1 + P_2 + \cdots + P_7 = 1$$

$$\cancel{0 + 2k + 3k}\; 0 + k + 2k + 2k + 3k + k^2 + 2k^2 + \cdots 7k^2 + k = 1$$

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

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

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

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

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

$$k = -1 \quad \Big| \quad 10k - 1 = 0, \quad k = 1/10$$

$$\therefore k = \frac{1}{10} \Rightarrow 0.1$$

The probability distribution is given by

x	0	1	2	3	4	5	6	7
$P(x)$	0	0.1	0.2	0.2	0.3	0.01	0.02	0.17

(ii) $P(1 \le x \le 7) = P(1) + P(2) + P(3) + P(4) + P(5) + P(6)$

$$= 0.1 + 0.2 + 0.3 + 0.01 + 0.02 + 0.2$$

$$P(1 \le x < 7) = 0.83$$

(iii) $P(x \le k) > \frac{1}{2}$

when $k = 0$, $P(x \le 0) = P(0) = 0 \not> \frac{1}{2}$

when $k = 1$, $P(x \le 1) = P(0) + P(1)$

$$= 0 + 0.1 = 0.1 \not> \frac{1}{2}$$

$k = 2$, $P(x \le 2) = P(0) + P(1) + P(0)$

$$= 0 + 0.1 + 0.2 = 0.3 \not> \frac{1}{2}$$

$k = 3$, $P(x \le 3) = P(0) + P(1) + P(2) + P(3)$

$$= 0 + 0.1 + 0.2 + 0.2 = 0.5 \not> \frac{1}{2}$$

$k = 4$, $P(x \le 4) = P(0) + P(1) + P(2) + P(3) + P(4)$

$$= 0 + 0.1 + 0.2 + 0.2 + 0.3 = 0.8 > \frac{1}{2}$$

$k = 5$, $P(x \le 5) = P(0) + \cdots + P(5)$

$$= 0 + 0.1 + 0.2 + 0.2 + 0.3 + 0.01 = 0.81 > \frac{1}{2}$$

$k = 6$, $P(x \le 6) = P(0) + \cdots P(6)$

$$= 0 + 0.1 + 0.2 + 0.2 + 0.3 + 0.01 + 0.02 = 0.83$$

Minimum

$K = 4 \quad \left(0.8 > \frac{1}{2}\right) - P(X \le K) > \frac{1}{2}$

(iv) Probability distribution function (PDF) $F_X(x) = P(X \le x)$

X	$F_X(x) = P(X \le x)$
when $X=0$	$F_X(0) = P(X \le 0) = P(0) = 0$
$X=1$	$F_X(1) = P(X \le 1) = P(0)+P(1) = 0+0.1 = 0.1$
$X=2$	$F_X(2) = P(X \le 2) = P(0)+P(1)+P(2) = 0+0.1+0.2 = 0.3$
$X=3$	$F_X(3) = P(X \le 3) = P(0)+P(1)+P(2)+P(3) = 0+0.1+0.2+0.2 = 0.5$
$X=4$	$F_X(4) = P(X \le 4) = P(0)+P(1)+P(2)+P(3)+P(4)$ $= 0+0.1+0.2+0.2+0.3 = 0.8$
$X=5$	$F_X(5) = P(X \le 5) = P(0)+P(1)+P(2)+P(3)+P(4)+P(5)$ $= 0+0.1+0.2+0.2+0.3+0.01 = 0.81$
$X=6$	$F_X(6) = P(X \le 6) = P(0)+P(1)+P(2)+P(3)+P(4)+P(5)+P(6)$ $= 0+0.1+0.2+0.2+0.3+0.01+0.02 = 0.83$
$X=7$	$F_X(7) = P(X \le 7) = P(0)+P(1)+P(2)+P(3)+P(4)+P(5)+P(6)+P(7)$ $= 0+0.1+0.2+0.2+0.3+0.01+0.02+0.17 = 1$

(v) Mean: $\mu = E(X) = \sum_{i=0}^{7} x_i P_i$

$= x_0P_0 + x_1P_1 + x_2P_2 + x_3P_3 + x_4P_4 + x_5P_5 + x_6P_6 + x_7P_7$

$= 0(0) + 1(0.1) + 2(0.2) + 3(0.2) + 4(0.3) + 5(0.01) + 6(0.02) + 7(0.17)$

$\boxed{\mu = E(X) = 3.66}$

(vi) variance : $\sigma^2 = E(X^2) =$

A random variable X has the function probability distribution.

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

(i) Determine k
(ii) Mean
(iii) Variance
(iv) Standard deviation.

3. Find the mean and variance of uniform probability distribution of a function F(X) is given by,

$$F(x) = \frac{1}{n}, \ x = 1, 2, 3, \dots, n.$$

Sol

The uniform prob. function of F(x) is

$$F(x) = \frac{1}{n}; \ x = 1, 2, 3, \dots, n.$$

X	1	2	3	.....	n
F(x)	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	.....	$1/n$

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

$$= x_1 P_1 + x_2 P_2 + \dots + x_n P_n$$

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

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

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

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

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

$$= \left[x_1^2 P_1 + x_2^2 P_2 + x_3^2 P_3 + \dots + x_n^2 P_n\right] - \mu^2$$

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

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

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

$$= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}$$

$$= \frac{2(2n^2+n+2n+1) - (n^2+1+2n)3}{12}$$

$$= \frac{4n^2+6n+2-3n^2-3-6n}{12}$$

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

4. Let X denotes the minimum of 2 numbers that appear when a pair of fair dice are thrown.

Determine:

(i) Discrete prob. distribution.

(ii) Exception

(iii) Variance

(iv) Standard deviation.

Sol Let X denotes the min of 2 numbers when 2 dice are thrown.

The no. of outcomes are 36 given by

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

for minimum $X = 1 \rightarrow \{(1,1), (1,2)\ (1,3)\ (1,4)\ (1,5)\ (1,6)\ (2,1)\ (3,$
$(4,1)\ (5,1)\ (6,1)\}$

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

for min $X = 2 \rightarrow \{(2,2)\ (2,3)\ (2,4)\ (2,5)\ (2,6)\ (3,2)\ (4,2)\ (5,2)\ (6,$

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

for min $X = 3 \rightarrow \{(3,3)\ (4,3)\ (5,3)\ (6,3)\ (3,4)\ (3,5)\ (3,6)$

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

for min $X = 4 \rightarrow \{(4,4)\ (4,5)\ (4,6)\ (5,4)\ (6,4)\}$

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

for min $X = 5 \Rightarrow \{(5,5)\ (5,6)\ (6,5)\}$

$$P(X=15) = 3/36$$

for min $X = 6 \Rightarrow \{(6,6)\}$

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

Probability distribution is given by

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

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

$$= x_1P_1 + x_2P_2 + x_3P_3 + x_4P_4 + x_5P_5 + x_6P_6$$

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

$$= \frac{11 + 18 + 21 + 20 + 15 + 6}{36} = \frac{91}{36}$$

$$\mu = 2.5278$$

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

$= x_1^2 P_1 + x_2^2 P_2 + x_3^2 P_3 + x_4^2 P_4 + x_5^2 P_5 + x_6^2 P_6 - \mu^2$

$= (1)^2\left(\frac{11}{36}\right) + 2^2\left(\frac{9}{36}\right) + 3^2\left(\frac{7}{36}\right) + 4^2\left(\frac{5}{36}\right) + 5^2\left(\frac{3}{36}\right) + 6^2\left(\frac{1}{36}\right) - \left(\frac{91}{36}\right)^2$

$= \frac{11+36+63+80+75+36}{36} - (2.5278)^2$

$= \frac{301}{36} - (2.5278)^2$

$= \frac{301}{36} - 6.3898 = \frac{301 - 230.0328}{36} = \frac{70.9682}{36}$

$\sigma^2 = 1.9713$

standard deviation:

$\sigma = \sqrt{1.9713} = 1.4040$

(2) Continuous Probability Distribution:

Consider an infinitely small interval $\left(x - \frac{dx}{2}, x + \frac{dx}{2}\right)$ of the length 'dx' around the point 'x'

Let $f(x)$ be any continuous of x so that $f(x)dx$ represents the probability that the variable 'x' falls in the infinitely small interval $\left(x - \frac{dx}{2}, x + \frac{dx}{2}\right)$

→ Symbolically it can represented as,

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

then $f(x)$ is called probability density function of a continuous random variable 'x'.

→ If the function $f(x)$ satisfies the 2 conditions

1. $f(x) \geq 0 \ \forall x$
2. $\int_{-\infty}^{+\infty} f(x)\,dx = 1$, then $f(x)$ is said to be probability mass function of X

* Mean of the continuous probability distribution is given by.

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

* Variance of the continuous probability distribution is given by

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

* Mean deviation of continuous probability distribution

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

* Median:

Median is the point which divides the entire distribution into 2 equal parts.

→ if X is defined from a to b & M is the median

then $\int_a^M f(x)dx = \int_M^b f(x)dx = \frac{1}{2}$, Solve for the value of M, we obtain the median.

* Mode:

Mode is the value of x for which f(x) is maximu[m] and is given by

$f'(x)=0$ &

$f''(x)<0$

Prob

① If the probability density function of a random variab[le] is given by $f(x)=\begin{cases} k(1-x^2) & \text{for } 0<x<1 \\ 0 & \text{for otherwise} \end{cases}$ find the value of k & the probabilities that the random variabl[e] will take on the values (i) b/w 0.1 & 0.2 (ii) >0.5

Sol

WKT

the probability function of a continuous random variable is $\int_{-\infty}^{\infty} f(x)dx = 1$

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

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

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

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

$$\Rightarrow k\left[1 - \frac{1}{3}\right] = 1$$

$$\Rightarrow k\left[\frac{2}{3}\right] = 1$$

$$\Rightarrow \boxed{k = \frac{3}{2}}$$

(i) $P(0.1 \le X \le 0.2)$

$$= \int_{0.1}^{0.2} f(x)dx$$

$$= \int_{0.1}^{0.2} \frac{3}{2}(1-x^2)dx$$

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

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

$$\cancel{= \frac{3}{2}\left((0.1973) - (0.09\right)}$$

$\cancel{(3107)}$

=

0.0417

② $P(X \geq 0.5) = \int_{0.5}^{\infty} f(x)\,dx$

$= \int_{0.5}^{1} f(x)\,dx + \int_{1}^{\infty} f(x)\,dx$

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

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

$= \frac{3}{2}\left[\frac{2}{3} - 0.4583\right]$

$= 0.31255 \simeq 0.3126$

② A Continuous random variable X is defined by

$$f(x) = \begin{cases} \frac{1}{16}(3+x)^2 \; ; & -3 \leq X \leq -1 \\ \frac{1}{16}(6-2x^2) \; ; & -1 \leq x \leq +1 \\ \frac{1}{16}(3-x)^2 \; ; & 1 \leq x \leq 3 \\ 0 & \text{elsewhere} \end{cases}$$

(i) Verify that $f(x)$ is a probability density function & also find mean of (x).

Sol WKT probability density function of a continuous random variable,

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

LHS:

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

$= \int_{-\infty}^{-3} f(x)\,dx + \int_{-1}^{-1} f(x)\,dx + \int f \int_{1}^{3} f(x)\,dx + \int_{3}^{\infty} f(x)\,dx$

$$= \{0 + \frac{1}{16}\left[\int_{-3}^{-1}(3+x)^2 + \int_{-1}^{1}(6-2x^2) + \int_{1}^{3}(3-x)^2\right]dx + 0$$

$$= \frac{1}{16}\left[\int_{-3}^{-1}(9+x^2+6x)dx + \int_{-1}^{1}(6-2x^2)dx + \int_{1}^{3}(9+x^2-6x)dx\right)$$

$$= \frac{1}{16}\left[\left(9x + \frac{x^3}{3} + \frac{6x^2}{2}\right)_{-3}^{-1} + \left(6x - \frac{2x^3}{3}\right)_{-1}^{1} + \left(9x + \frac{x^3}{3} - \frac{6x^2}{2}\right)_{1}^{3}\right]$$

$$= \frac{1}{16}\left[\left(9(-1) + \frac{(-1)^3}{3} + \frac{6}{2}\right) - \left(9(-3) + \frac{(-3)^3}{3} + \frac{54}{2}\right)\right.$$

$$+ \left(6 - \frac{2}{3}\right) - \left(-6 + \frac{2}{3}\right)$$

$$\left. + \left(9(3) + \frac{3^3}{3} - \frac{6(3)^2}{2}\right) - \left(9 + \frac{1}{3} - \frac{6}{2}\right)\right)$$

$$= \frac{1}{16}\left(\left(-9 - \frac{1}{3} + \frac{6}{2}\right) + \cancel{27} + 9 \cancel{27} + \frac{16}{3} \cancel{+6-\frac{2}{3}} + \frac{16}{3}\right.$$

$$\left. + \cancel{27} + \cancel{27}\,9 - \cancel{27} - \cancel{9} - \frac{1}{3} + 3\right)$$

$$= \frac{1}{16}\left(-6 - \frac{1}{3}\right\} + \cancel{54} + \frac{32}{3} + 9 - \frac{1}{3} + 3$$

$$= \frac{1}{16}\left(6 - \frac{2}{3} + \cancel{54} + \frac{32}{3}\right)$$

$$= \frac{1}{16}\left(\frac{18 - 2 + 32}{3}\right) = \frac{48}{48} = 1$$

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

Hence proved

(ii) Mean

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

$$= \int_{-\infty}^{-3} x f(x)\,dx + \int_{-3}^{-1} x f(x)\,dx + \int_{-1}^{1} x f(x)\,dx + \int_{1}^{3} x f(x)\,dx + \int_{3}^{\infty} x f(x)\,dx$$

$$= 0 + \int_{-3}^{-1} \frac{1}{16} x (3+x)^2\,dx + \int_{-1}^{1} \frac{1}{16} x (6-2x^2)\,dx + \int_{1}^{3} \frac{1}{16} x (3-x)^2\,dx$$

$$= \frac{1}{16}\left[\int_{-3}^{-1} x(9+x^2+6x)\,dx + \int_{-1}^{1}(6x-2x^3)\,dx + \int_{1}^{3}(9+x^2-6x)x\,dx\right]$$

$$= \frac{1}{16}\left[\int_{-3}^{-1}(9x+x^3+6x^2)\,dx + \int_{-1}^{1}(6x-2x^3)\,dx + \int_{1}^{3}(9x+x^3-6x^2)\,dx\right]$$

$$= \frac{1}{16}\left(\left[\frac{9x^2}{2}+\frac{x^4}{4}+\frac{6x^3}{3}\right]_{-3}^{-1} + \left[\frac{6x^2}{2}-\frac{2x^4}{4}\right]_{-1}^{1} + \left[\frac{9x^2}{2}+\frac{x^4}{4}-\frac{6x^3}{3}\right]_{1}^{3}\right)$$

$$= \frac{1}{16}\left(\left(\frac{9}{2}+\frac{1}{4}-\frac{6}{3}\right)-\left(\frac{81}{2}+\frac{6561}{4}-\frac{4374}{3}\right)+\left(3-\frac{1}{2}\right)-\left(3-\frac{1}{2}\right) + \left(\frac{81}{2}+\frac{6541}{4}-\frac{4374}{3}\right)-\left(\frac{9}{2}+\frac{1}{4}-\frac{6}{3}\right)\right)$$

$$= \frac{1}{16}(0)$$

$$= 0$$

$$\therefore \text{Mean} = \int_{-\infty}^{\infty} x f(x)\,dx = 0$$

③ For a continuous random variable x whose probability distribution function is given by

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

find the value of, c, mean and variance of X.

Sol $f(x) =$

④ A continuous random variable X has a probability density function $f(x) = \begin{cases} k \cdot x e^{-\lambda x} ; & \cancel{0 \le x \le 1} \ x \ge 0, \lambda > 0 \\ 0 ; & \text{elsewhere.} \end{cases}$ determine 'k value, mean & variance of X.

Sol WKT

the probability distribution function

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

$$= \cancel{\int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{\infty} f(x)dx}$$

$$\cancel{= 0 + \int_{0}^{1} k x e^{-\lambda x}}$$

$$\cancel{= k\int_{0}^{1} x e^{-\lambda x} = 1}$$

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

$$\Rightarrow 0 + \int_{0}^{\infty} k \cdot x e^{-\lambda x} = 1$$

$$\Rightarrow k\left[x \cdot \frac{e^{-\lambda x}}{-\lambda} - 1 \cdot \frac{e^{-\lambda x}}{\lambda^2}\right]_0^{\infty} = 1$$

$$\Rightarrow k\left[(0-0) - \left(0 - \frac{1}{\lambda^2}\right)\right] = 1$$

$$\Rightarrow k\left[\frac{1}{\lambda^2}\right] = 1 \quad \Rightarrow \boxed{k = \lambda^2}$$

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

$= \int_{-\infty}^{0} x f(x) dx + \int_{0}^{\infty} x f(x) dx.$

$= 0 + \lambda^2 \int_0^\infty x \cdot x e^{-\lambda x} dx$

$= \lambda^2 \int_0^\infty x^2 e^{-\lambda x} dx$

$= \lambda^2 \cdot \left[x^2 \cdot \frac{e^{-\lambda x}}{-\lambda} - 2x \cdot \frac{e^{-\lambda x}}{\lambda^2} + 2 \cdot \frac{e^{-\lambda x}}{-\lambda^3} \right]_0^\infty$

$= \lambda^2 \left[(0-0+0) - \left(0 - 0 + \frac{2(1)}{-\lambda^3}\right)\right]$

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

$\boxed{\mu = \frac{2}{\lambda}}$

Variance:

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

$= \int_{-\infty}^{0} x^2 f(x) dx + \int_0^\infty x^2 f(x) dx - \mu^2$

$= 0 + \lambda^2 \int_0^\infty x^3 \cdot e^{-\lambda x} dx - \mu^2$

$= \lambda^2 \left[x^3 \cdot \frac{e^{-\lambda x}}{-\lambda} - 3x^2 \frac{e^{-\lambda x}}{\lambda^2} + 6x \frac{e^{-\lambda x}}{-\lambda^3} - 6 \cdot \frac{e^{-\lambda x}}{\lambda^4} \right]_0^\infty - \mu^2$

$= \lambda^2 \left[(0-0+0-0) - \left(0 - 0 + 0 - 6\left(\frac{1}{\lambda^4}\right)\right] - \left(\frac{2}{\lambda}\right)^2$

$= \lambda^2 \left(\frac{6}{\lambda^4}\right) - \frac{4}{\lambda^2}$

$\sigma^2 = \frac{6}{\lambda^2} - \frac{4}{\lambda^2}$ $\therefore \boxed{\text{Variance} = \frac{2}{\lambda^2}}$

12/11/22

Unit - II

Expectation and Discrete Probability Distribution

There are various Probability Distribution given by

i, Bernoll's Distribution

ii, Binomial Distribution

iii, Poission Distribution

iv, Rectangular Distribution

v, Negative Binomial Distribution

vi, Geometric Distribution

Bernoull's Distribution:

A random variable x which takes the values 0 and 1 with respective Probabilities q, p i.e., $P(x=0) = q$, $P(x=1) = p$ and $p+q=1$ is called a Bernoull's Discrete Random variable and is said to have Bernoull's Distribution

→ The Probability Density function of Bernoull's Distribution is given by

$$P(X=x) = p(x) = p^x q^{(1-x)} \; ; \; x = 0, 1$$

→ Mean of the Bernoull's Distribution is given by

$$\mu = \sum_{i=0}^{1} x_i p_i$$

$$= x_0 p_0 + x_1 p_1 = 0(q) + 1(p) = 0 + p \quad \boxed{\mu = p}$$

Scanned with OKEN Scanner

→ Variance of the Bernoulli's Distribution is given

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

$$= \sum_{i=0}^{1} x_i^2 p_i - \mu^2$$

$$= x_0^2 p_0 + x_1^2 p_1 - \mu^2$$

$$= 0^2(q) + 1^2(p) - p^2$$

$$= 0 + p - p^2$$

$$= p(1-p)$$

$$\boxed{\sigma^2 = pq}$$

→ Standard Deviation of the Bernoulli's Distribution is given by $S.D(\sigma) = \sqrt{variance}$

$$\therefore \boxed{\sigma = \sqrt{pq}}$$

Binomial Distribution:-

A random variable x defined for non negative values of x, the Probability Distribution function for Binomial Distribution is given by

$P(X=x) = P(x) = {}^nC_x\, p^x q^{n-x}$ for $x \geq 0$ & 0 for otherwise

also we define.

$P(X=r) = p(r) = {}^nC_r\, p^r q^{n-r}$; $r \geq 0$

→ Mean of the binomial distribution is given by

Mean $\mu = \sum_{r=0}^{n} r p(r) \Rightarrow \sum_{r=0}^{n} r p(r)$

$$\mu = \sum_{r=0}^{n} r \cdot {}^nC_r p^r q^{n-r}$$

$$= 0 + 1 \cdot {}^nC_1 p^1 q^{n-1} + 2\,{}^nC_2 p^2 q^{n-2} + \cdots + n \cdot {}^nC_n p^n q^0$$

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

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

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

$$= np(1)^{n-1}$$

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

→ Variance of the binomial Distribution is given by

$$\text{Variance } (\sigma^2) = E(x^2) - (E(x))^2$$

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

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

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

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

$$= 0 + 0 + 2\,{}^nC_2 p^2 q^{n-2} + 6\,{}^nC_3 p^3 q^{n-3} + \cdots$$

$$\cdots + n(n-1)\,{}^nC_n p^n q^0 + \mu - \mu^2$$

$$= 2 \cdot \frac{n(n-1)}{2!} p^2 q^{n-2} + 6 \cdot \frac{n(n-1)(n-2)}{3!} p^3 q^{n-3} + \dots +$$

$$n(n-1) 1 \cdot p^n + \mu - \mu^2$$

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

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

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

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

$$= np(1-p)$$

$$= npq$$

$$\boxed{\sigma^2 = npq}$$

standard Deviation of the Binomial Distribution is given by

$$S.D(\sigma) = \sqrt{\text{variance}}$$

$$\boxed{\sigma = \sqrt{npq}}$$

Recurrence of Binomial Distribution

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

$$P(r+1) = {}^nC_{r+1} \, p^{r+1} \, q^{n-(r+1)}$$

$$\frac{P(r+1)}{P(r)} = \frac{{}^nC_{r+1}}{{}^nC_r} \cdot \frac{p^{r+1}}{p^r} \cdot \frac{q^{n-r-1}}{q^{n-r}}$$

$$= \frac{{}^nC_{r+1}}{{}^nC_r} \cdot \frac{\cancel{p^r} \cdot p}{\cancel{p^r}} \cdot \frac{\cancel{q^{n-r}} \cdot q^{-1}}{\cancel{q^{n-r}}}$$

$$\frac{P(r+1)}{P(r)} = \left(\frac{n-r}{r+1}\right)\frac{p}{q}$$

Binomial frequency distribution is given by

$$P(r+1) = \left(\frac{n-r}{r+1}\right)\frac{p}{q} \cdot P(r)$$

$$\boxed{N(q+p)^n}$$

where, N = Sum of frequencies

q = Probability of failure

p = Probability of success

n = No. of trails

Poission's Distribution:-

Poission's Distribution is the limiting case of Binomial Distribution in the condition

i, n is very large

ii, p is very small and

Product of n and p equal to λ (constant)

→ The Random variable x which is defined only for non-negative values of x, whose Probability distribution function is given by

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

→ Mean of Poission's Distribution is given by

$$\text{Mean, } E(x) \text{ or } \mu = \sum_{x=0}^{n} x \cdot p(x)$$

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

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

= Put $x-1 = y \Rightarrow x = y+1$

$$= \sum_{y=0}^{n} \frac{e^{-\lambda} \cdot \lambda^{y+1}}{y!}$$

$$= e^{-\lambda} \sum_{y=0}^{n} \frac{\lambda^y \cdot \lambda^1}{y!}$$

$$= e^{-\lambda} \cdot \lambda \sum_{y=0}^{n} \frac{\lambda^y}{y!}$$

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

$$= \lambda$$

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

→ Variance of Poission's Distribution is given by

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

$$= \sum_{x=0}^{n} x^2 p(x) - \mu^2$$

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

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

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

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

Put $x-2=y$ $\quad x-1=z$

$x=y+2$ $\quad x=z+1$

$$= e^{-\lambda} \sum_{y=0}^{n} \frac{\lambda^{y+2}}{y!} + e^{-\lambda} \sum_{z=0}^{n} \frac{\lambda^{z+1}}{z!} - \lambda^{2}$$

$$= e^{-\lambda} \sum_{y=0}^{n} \frac{\lambda^{y} \cdot \lambda^{2}}{y!} + e^{-\lambda} \sum_{z=0}^{n} \frac{\lambda^{z} \cdot \lambda^{1}}{z!} - \lambda^{2}$$

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

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

$$\boxed{\sigma^{2} = \lambda}$$

Standard Deviation of Poisson's Distribution is give by $S.D\ (\sigma) = \sqrt{\text{variance}}$

$$\boxed{\sigma = \sqrt{\lambda}}$$

Recurrence of Poission's Distribution

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

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

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

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

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

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

Poission's frequency Distribution is given by

$$\boxed{N \cdot P(x)}$$

Q) 10 coins are tossed simultaneously find the Probability that getting

i) Atleast 7 heads

ii) Atleast 5 heads

iii) Atmost 3 heads

Given that,

No. of coins (n) = 10

Probability of getting head (p) = $\frac{1}{2}$

Probability of Not getting head (q) = $1-P = \frac{1}{2}$

Since Probability of success, failure and no. of trails are small so we use Binomial Distribution.

i) $P(x \geq 7) = P(x=7) + P(x=8) + P(x=9) + P(x=10)$

$$= {}^{10}C_7 (0.5)^7 \cdot (0.5)^{10-7} + {}^{10}C_8 (0.5)^8 \cdot (0.5)^2 + {}^{10}C_9 (0.5)^9 (0.5)^1 + {}^{10}C_{10} (0.5)^{10} \cdot (0.5)^0$$

$(= 0.1179 + 0.0439)$

$$= (0.5)^{10} \left[{}^{10}C_7 + {}^{10}C_8 + {}^{10}C_9 + {}^{10}C_{10} \right]$$

$= 0.1718 \sim 0.1719$

ii) $P(x \geq 5) = P(x=5) + P(x=6) + P(x=7) + P(x=8) + P(x=9) + P(x=10)$

$$= {}^{10}C_5 (0.5)^5 \cdot (0.5)^5 + {}^{10}C_6 (0.5)^6 \cdot (0.5)^4 + {}^{10}C_7 (0.5)^7 \cdot (0.5)^3 + {}^{10}C_8 (0.5)^8 \cdot (0.5)^2 + {}^{10}C_9 (0.5)^9 (0.5)^1 + {}^{10}C_{10} (0.5)^{10} \cdot (0.5)^0$$

$$= (0.5)^{10} \left[{}^{10}C_5 + {}^{10}C_6 + {}^{10}C_7 + {}^{10}C_8 + {}^{10}C_9 + {}^{10}C_{10} \right]$$

$= 0.6230$

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

$$= {}^{10}C_3 (0.5)^3 (0.5)^7 + {}^{10}C_2 (0.5)^2 (0.5)^8 + {}^{10}C_1 (0.5)^1 (0.5)^9$$

$$= (0.5)^{10} \left[{}^{10}C_3 + {}^{10}C_2 + {}^{10}C_1 \right]$$

$$= 0.1709$$

Q) 20% of the items are produced from a factory are defective, find probability that in a sample of 5 choosen at random

i, None is defective

ii, only 1 is defective

iii, $P(1 < x < 4)$

Given that,

Probability of success $(p) = 20\% = 0.2$

$n = 5$

Probability of failure $(q) = 1 - p = 1 - 0.2 = 0.8$

Since, n is small and probability of success, failures are available we use binomial distribution

Probability function of Binomial distribution is given by

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

i, $P(\text{None}) = P(x=0)$

$$= {}^5C_0 (0.2)^0 (0.8)^5$$

$$P(x=0) = 0.3276$$

ii, $P(1) = P(x=1)$

$$= {}^5C_1 (0.2)^1 (0.8)^4$$

$$P(x=1) = 0.4096$$

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

$$= {}^5C_2 (0.2)^2 (0.8)^3 + {}^5C_3 (0.2)^3 (0.8)^2$$

$$P(1<x<4) = 0.256$$

ii) Assume that 50% of all Engineering students are good in Mathematics determine the Probability that among 18 Engineering students

i, Exactly 10 ii, Atleast 10 iii, Atmost 8

iv, Atleast 2 & Atmost 9 are good in mathematics

Given that,

Probability of success $(p) = 50\% = 0.5$

$n = 18$

Probability of failure $(q) = 1 - p = 1 - 0.5 = 0.5$

Since, Probability of success and failure, $n=18$ available we use Binomial distribution.

i, $P(X=10) = {}^{18}C_{10} \cdot (0.5)^{10} (0.5)^{8}$

$= 0.1669$

ii, $P(X \geq 10) = P(X=10) + P(X=11) + P(X=12) + P(X=13) +$
$P(X=14) + P(X=15) + P(X=16) + P(X=17) + P(X=18)$

$= {}^{18}C_{10} (0.5)^{10} (0.5)^{8} + {}^{18}C_{11} (0.5)^{11} (0.5)^{7} + {}^{18}C_{12} (0.5)^{12} (0.5)^{6} +$
${}^{18}C_{13} (0.5)^{13} (0.5)^{5} + {}^{18}C_{14} (0.5)^{14} (0.5)^{4} + {}^{18}C_{15} (0.5)^{15} (0.5)^{3} +$
${}^{18}C_{16} (0.5)^{16} (0.5)^{2} + {}^{18}C_{17} (0.5)^{17} (0.5)^{1} + {}^{18}C_{18} (0.5)^{18} (0.5)^{0}$

$=$ [illegible] $+$

$= (0.5)^{18} \left[{}^{18}C_{10} + {}^{18}C_{11} + {}^{18}C_{12} + {}^{18}C_{13} + {}^{18}C_{14} + {}^{18}C_{15} + {}^{18}C_{16} + {}^{18}C_{17} + {}^{18}C_{18} \right]$

$= 0.4072$

iii, $P(x \leq 8) = P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=7)$
$+ P(X=8)$

$= {}^{18}C_{1} (0.5)^{1} (0.5)^{17} + {}^{18}C_{2} (0.5)^{2} (0.5)^{16} + {}^{18}C_{3} (0.5)^{3} (0.5)^{15} +$
${}^{18}C_{4} (0.5)^{4} (0.5)^{14} + {}^{18}C_{5} (0.5)^{5} (0.5)^{13} + {}^{18}C_{6} (0.5)^{6} (0.5)^{12} +$
${}^{18}C_{7} (0.5)^{7} (0.5)^{11} + {}^{18}C_{8} (0.5)^{8} (0.5)^{10}$

$$= (0.5)^{18}\left[{}^{18}C_1 + {}^{18}C_2 + {}^{18}C_3 + {}^{18}C_4 + {}^{18}C_5 + {}^{18}C_6 + {}^{18}C_7 + {}^{18}C_8\right]$$

$$= 0.4073$$

iv, $P(2 \leq x \leq 9) = P(x=2) + P(x=3) + P(x=4) + P(x=5) + P(x=6)$

$$P(x=7) + P(x=8) + P(x=9)$$

$$= {}^{18}C_2 (0.5)^2 (0.5)^{16} + {}^{18}C_3 (0.5)^3 (0.5)^{15} + {}^{18}C_4 (0.5)^4 (0.5)^{14} +$$

$${}^{18}C_5 (0.5)^5 (0.5)^{13} + {}^{18}C_6 (0.5)^6 (0.5)^{12} + {}^{18}C_7 (0.5)^7 (0.5)^{11}$$

$$+ {}^{18}C_8 (0.5)^8 (0.5)^{10} + {}^{18}C_9 (0.5)^9 (0.5)^9$$

$$= (0.5)^{18}\left[{}^{18}C_2 + {}^{18}C_3 + {}^{18}C_4 + {}^{18}C_5 + {}^{18}C_6 + {}^{18}C_7 + {}^{18}C_8 + {}^{18}C_9\right]$$

$$= 0.5926.$$

Q) In a Binomial distribution consisting of 5 Independent trails, Probability of 1 & 2 success are $P(x=1) = 0.4096$ $P(x=2) = 0.2048$. Find the Parameter distribution

Given that,

$n=5$

$P(x=1) = 0.4096$ — ① $\quad P(x=2) = 0.2048$ — ②

$\therefore \frac{1}{2}$

$$\frac{P(x=1)}{P(x=2)} = \frac{0.4096}{0.2048} = 2$$

0.5426

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

$$\frac{{}^5C_1\, p^1 q^4}{{}^5C_2\, p^2 q^3} = \frac{2}{1}$$

$$\frac{1\,\cancel{5}q}{\cancel{10}_2\, p} = \frac{2}{1}$$

$$\frac{1-p}{2p} = \frac{2}{1} \qquad 1-p = q$$

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

$$\boxed{p = \frac{1}{5}} \qquad \boxed{q = \frac{4}{5}}$$

Q) The Mean and variance of Binomial Distribution are 16 and 8 respectively. then determine

i) $P(x \geq 1)$ ii) $P(x > 2)$

Given,

Mean of the Binomial Distribution

i.e., $\mu = np = 16$ —(1)

variance of the Binomial Distribution

$\sigma^2 = npq = 8$ —(2)

$$\frac{(2)}{(1)} = \frac{npq}{np} = \frac{8}{16} \qquad q = 1/2 \qquad p = \frac{1}{2}$$

from 1, $np = 16 \Rightarrow$

$\boxed{n = 32}$

i) $P(x \geq 1) = 1 - P(x < 1)$

$= 1 - P(0)$

$= 1 - {}^{n}C_0 \, p^0 q^{n-0}$

$= 1 - {}^{32}C_0 \, p^{(0.5)} q^{32}$

$=$

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

$= 1 - [{}^{32}C_0 \, P(0.5)^0 (0.5)^0 + {}^{32}C_1 (0.5)^1 (0.5)^{32} + {}^{32}C_2 (0.5)$

(0.5)

Q) The Mean and variance of Binomial Distribution are 3 & 9

i) value of n ii) $P(x \geq 7)$ iii) $P(1 \leq x \leq 6)$

Q) Fit a Binomial Distribution to the following frequency Distribution table

x	0	1	2	3	4	5
f	2	14	20	34	22	8

Q) Fit a Binomial Distribution to data

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

Binomial frequency distribution is given by

$$N(q+p)^n$$

where; n = no. of trials = 6

$N = 200$ = Sum of frequencies

Mean, $\mu = \frac{\sum f_i x_i}{\sum f_i} \Rightarrow \mu = \frac{535}{200} = 2.675$

Mean of B.D $\Rightarrow \mu = np$

$$2.675 = 6 \times P$$

$$\boxed{P = 0.4458}$$

$$\boxed{q = 1-p = 0.5542}$$

If B. frequency Distribution given by

~~N (q+q)~~ $N(q+p)^n$

$$= 200\left[0.5542 + 0.4458\right]^6$$

$= 200\Big[{}^6C_0 (0.5542)^0 (0.4458)^6 + {}^6C_1 (0.5542)^1 (0.4458)^5 +$

${}^6C_2 (0.5542)^2 (0.4458)^4 + {}^6C_3 (0.5542)^3 (0.4458)^3 +$

${}^6C_4 (0.5542)^4 (0.4458)^2 + {}^6C_5 (0.5542)^5 (0.4458)^1 +$

${}^6C_6 (0.5542)^6 (0.4458)^0$

$= 200(0.0078 + 0.0585 + 0.1819 + 0.3016 + 0.2812 + 0.1396 +$

0.0290

$= 1.56 + 11.9097 + 36.38 + 60.32 + 56.24 + 27.96 + 5.$

$= 2 + 12 + 36 + 60 + 56 + 28 + 6$

Theoritical frequencies are given by

x	0	1	2	3	4	5	6
f	13	52	~~32~~ 58	38	32	16	4
	2	12	36	60	56	28	6

Q)

x	0	1	2	3	4	5
f	2	14	20	34	22	8

Binomial frequencies distribution is given by

$N(q+p)^n$

where $n = 5$

$N = 100$

Mean $\mu = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{284}{100} = 2.84$

Mean of B.D $= \mu = np$

$2.84 = 5 \times p$

$\boxed{P = 0.568}$

$q = 1 - p = 1 - 0.568 = 0.432$

$\boxed{q = 0.432}$

Binomial Distribution given by

$N(q+p)^n$

$100(0.432+0.568)^5$

$= 100[{}^5C_0 (0.432)^0 (0.568)^5 + {}^5C_1 (0.432)^1 (0.568)^4 +$
${}^5C_2 (0.432)^2 (0.568)^3 + {}^5C_3 (0.432)^3 (0.568)^2 +$
${}^5C_4 (0.432)^4 (0.568)^1 + {}^5C_5 (0.432)^5 (0.568)^0)$

$= 100[0.0591 + 0.2248 + 0.3420 + 0.2601 + 0.0989$
$+ 0.0150]$

$= 5.91 + 22.48 + 34.2 + 26.01 + 9.89 + 1.5$

$= 6 + 22 + 34 + 26 + 10 + 2$

Theoritical frequencies are given by

x	0	1	2	3	4	5
f	2	14	20	34	22	8
$Bf(x)$	6	22	34	26	10	2

Q) Given B.D's are 3 & 9/4

Mean of the B.P, $\mu = np$

$3 = np$ — (1)

Variance $\sigma^2 = npq$

$npq = 9/4$ — (2)

$$\frac{2}{1} \Rightarrow \frac{\cancel{n}pq}{\cancel{n}p} = \frac{9/4}{3} = \frac{9}{12} = \frac{3}{4} = 0.75$$

$$q = \frac{3}{4} = 0.75$$

$$p = 1 - q = 1 - 0.75$$

$$p = 0.25$$

i, from ① $\rightarrow np = 3$.

$$n(0.25) = 3$$

$$\boxed{n = 12}$$

ii, $P(x \geq 7) = 1 - P(x < 7)$

$$= 1 - [P(0) + P(1) + P(2) + P(3) + P(4) + P(5) + P(6) + P(7)]$$

$$= 1 - [{}^{12}C_0 p^0 q^{12} + {}^{12}C_1 (0.25)^1 (0.75)^{11} + {}^{12}C_2 (0.25)^2 (0.75)^{10} +$$

$${}^{12}C_3 (0.25)^3 (0.75)^9 + {}^{12}C_4 (0.25)^4 (0.75)^8 + {}^{12}C_5 (0.25)^5 (0.75)$$

$${}^{12}C_6 (0.25)^6 (0.75)^6 + {}^{12}C_7 (0.25)^7 (0.75)^5]$$

$$= 1 - 0.9972 = 0.0028$$

iii, $P(1 \leq x < 6) = 1 - [P(1) + P(2) + P(3) + P(4) + P(5)]$

$$= 1 - [{}^{12}C_1 p^1 q^{11} + {}^{12}C_2 p^2 q^{10} + {}^{12}C_3 p^3 q^9 + {}^{12}C_4 p^4 q^8 + {}^{12}C_5 p^5 q^7]$$

$$= 1 - 0.9139$$

$$= 0.0861$$

1Q) If the Probability that an individual suffers a Bad reaction from a certain injection is 0.001, determine the Probability that out of 2000 individuals

i, Exactly 3 ii, more than 2 iii, None

iv, More than one individual suffer a bad reaction

2Q) If 2% of items of a factory are defective, the items are Packed in boxes then what is the Probability that there will be

i, 2 Defective items ii, Atleast 3 Defective items in a box of 100 items

3Q) The Avg no. of accidents on any day on National highway is 1.6. Determine the Probability that No. of Accidents are

i, Atleast 1 ii, Atmost 1 iii, No Accidents

4Q) The variance of Poison variant is 3 then find the Probability that

i, At $x=0$ ii, $0<x\leq 3$ iii, $1\leq x<4$

5Q) If x is Poison variant such that $\frac{3}{2}P(x=1) = P(x=3)$ then find $P(x\geq 1)$; $P(x\leq 3)$; $P(2\leq x\leq 5)$

i, ii, iii,

6Q) If X is a Bi[illegible]

P(X=0) = P(X=1), find P(0) & using Recurrence relation find the Probabilities at $x = 1, 2, 3, 4, 5$

7Q) using Recurrence formula find Probabilities when $x = 0, 1, 2, 3, 4, 5$ If the Mean of Poision distribution is

8Q) Fit a Poison Distribution to the data

x	0	1	2	3	4	5
$f(x)$	42	33	14	6	4	1

9Q) Find the Excepted frequencies using Poison Distribution,

x	0	1	2	3	4
$f(x)$	109	65	22	3	1

Sol:-

Poison frequency Distribution is given by

$N \cdot P(x)$

$N = 200$

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

Mean of Distribution:-

$$\mu = \frac{\sum f_i x_i}{\sum f_i}$$

$$\boxed{\mu = 0.61}$$

Mean of Poison Distribution

$\mu = \lambda = 0.61$

Poison frequencies are

$$N \cdot P(x) = N \cdot \left[\frac{e^{-\lambda} \cdot \lambda^{x}}{x!}\right]$$

when $x=0 \Rightarrow 200 \cdot \left[\frac{e^{-0.61} \cdot (0.61)^0}{0!}\right] = 108.6701 \simeq 108$

when $x=1 \Rightarrow 200 \cdot \left[\frac{e^{-0.61} \cdot (0.61)^1}{1!}\right] = 66.288 \simeq 66$

when $x=2 \Rightarrow 200 \cdot \left[\frac{e^{-0.61} \cdot (0.61)^2}{2!}\right] = 20.2180 \simeq 20$

when $x=3 \Rightarrow 200 \cdot \left[\frac{e^{-0.61} \cdot (0.61)^3}{3!}\right] = 4.1110 \simeq 4$

when $x=4 \Rightarrow 200 \cdot \left[\frac{e^{-0.61} \cdot (0.61)^4}{4!}\right] = 0.629 \simeq 1$

Theoritical frequencies are given by

x	0	1	2	3	4
$f(x)$	109	65	22	3	1
$P.f(x)$	108	66	20	4	1

19) Given that,

Probability of success $(p) = 0.001$

$n = 2000$

Since, n is large and P is very small we use Poison Distribution

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

$\lambda = n \cdot p$

$= 2000 \cdot 0.001$

$= 2$

i) P(Exactly 3) = $P(X=3)$

$$P(x) = \frac{e^{-2} \cdot 2^{3}}{3!} \doteq 0.1804$$

ii) $P(x \geq 2) \Rightarrow 1 - P(x \leq 2)$

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

$$= 1 - [0.1353 + 0.2707 + 0.2707$$

$$= 1 - 0.6767$$

$$= 0.3233$$

iii) P(None) = $P(X=0)$

$$= \frac{e^{-\lambda} \cdot \lambda^{x}}{x!} \Rightarrow \frac{e^{-2} \cdot 2^{0}}{0!}$$

$$= 0.1353$$

2(iv) P(more than 1) = $1 - P(x<1)$

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

$$= 1 - 0.1353$$

$$= 0.8647$$

28) Given that,

Probability of success $(p) = 0.02$

$n = 100$

Since P is small and n is large so, we use Poison Distribution

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

$\lambda = n \cdot p$

$= 100 \cdot 0.02$

$$\boxed{\lambda = 2}$$

i) P(2 Defective items) = $P(x=2)$

$$= \frac{e^{-2} \cdot 2^2}{2!}$$

$$= 0.2707$$

ii) P(Atleast 3) = $1 - P(x<3)$

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

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

$$= 1 - 0.6767 \Rightarrow 0.3233$$

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

3Q) Given that;

[Probability of success =] x

Mean No. of accidents $\lambda = 1.6 = \mu$

Mean of Poison Distribution $\mu = \lambda = 1.6$

i) P(Atleast 1) = $P(x \geq 1)$

$= 1 - P(x < 1)$

$= 1 - P(0)$

$= 1 - \frac{e^{-\lambda} \cdot \lambda^x}{x!} \Rightarrow 1 - \frac{e^{-1.6} \cdot (1.6)^0}{0!}$

$= 0.7981$

ii) P(Atmost 1) = $P(x \leq 1)$

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

$= \frac{e^{-1.6} \cdot (1.6)^0}{0!} + \frac{e^{-1.6} \cdot (1.6)^1}{1!}$

$= 0.2019 + 0.3230$

$= 0.5249$

5Q) Given that,

x is a Poison Variant

we have $\frac{3}{2} P(X=1) = P(X=3)$

$$\frac{3}{2} \cdot \frac{e^{-\lambda} \cdot \lambda^{1}}{1!} = \frac{e^{-\lambda} \cdot \lambda^{3}}{3!}$$

$$\lambda = \pm 3$$

since, the distribution is +ve $\therefore \boxed{\lambda = 3}$

i, $P(x \geq 1) = 1 - P(x < 1)$

$$= 1 - P(0)$$

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

$$=$$

ii, $P(x \leq 3) = P(0) + P(1) + P(2) + P(3)$

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

$$= 0.0498 + 0.1494 + 0.2240 + 0.4480$$

$$=$$

iii, $P(2 \leq x \leq 5) = P(2) + P(3) + P(4) + P(5)$

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

$$=$$

6Q) Given that,

x is a Poison variat

we have $P(x=0) = P(x=1)$

$$\frac{e^{-\lambda}\cdot\lambda^{0}}{0!} = \frac{e^{-\lambda}\cdot\lambda^{1}}{1!}$$

$$\boxed{\lambda=1}$$

$$P(0) = \frac{e^{-1}\cdot 1^{0}}{0!} = 0.3679$$

Recurrence Relation of Poison Distribution

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

when $x=0 \Rightarrow P(0+1) = \frac{1}{0+1}\cdot P(0)$

$P(1)$

$$= 1\times 0.3679$$

$$= 0.3679$$

$x=1 \Rightarrow P(2) = \frac{1}{1+1}\cdot P(1)$

$$= \frac{1}{2}\cdot(0.3679)$$

$$= 0.1839$$

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

$$= \frac{1}{3}(0.1839)$$

$$= 0.0613$$

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

$$= \frac{1}{4}(0.0613)$$

$$= 0.0153.$$

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

$= \frac{1}{5} (0.0153)$

$= 0.00306$

$x=5 \Rightarrow P(5+1) = \frac{1}{5+1} P(5)$

$\downarrow$

$P(6) = \frac{1}{6} (0.00306)$

$\simeq 0.00051$

78) Given

X is Poison variant

Mean of P.D $(\lambda) = 3$

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

Recurrence Relation of P.D

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

$x=0 \Rightarrow P(0+1) = \frac{1}{0+1} \cdot P(0)$

$= 1 \times 0.0498$

$= 0.0498$

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

$= \frac{1}{2} (0.0498) = 0.0249$

$x=2 \Rightarrow P(2+1) = \frac{1}{2+1} \cdot P(2)$

$= \frac{1}{3} (0.0249)$

Unit-3

Continuous Probability Distributions

There are various continuous probability distributions given by,

(i) Normal Distribution

(ii) Students, t-distribution

(iii) Snedecors f-distribution

(iv) Schi-square 'χ^2' Distribution

Normal distribution:

If x is a Random variable, is said to have a normal distribution if its density function are probability distribution function is given by

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; -\infty \le x \le \infty$$
$$-\infty \le \mu \le \infty$$
$$\sigma > 0$$

Constantles of Normal distribution

1. Mean of normal distribution:

Consider the normal distribution with b, σ as the parameter.

Then $f(x; b, \sigma) = \dfrac{e^{-\frac{(x-b)^2}{2\sigma^2}}}{\sigma.\sqrt{2\pi}}$

The mean $\mu = E(X)$ is given by

$$\mu = \int_{-\infty}^{\infty} x f(x) dx = \frac{1}{\sigma.\sqrt{2\pi}} \int_{-\infty}^{\infty} x . e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}\right)^2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + b) e^{-\frac{z^2}{2}} dz \quad \left[\text{Putting } z = \frac{x-b}{\sigma} \text{ so that } dz = \frac{dx}{\sigma}\right]$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{\frac{-z^2}{2}} dz + \frac{b}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-z^2}{2}} dz$$

$$= 0 + \frac{b}{\sqrt{2\pi}} \cdot 2 \int_0^\infty e^{-\frac{z^2}{2}} dz \left[\because z e^{-\frac{z^2}{2}} \text{ is odd function \& } e^{-\frac{z^2}{2}} \text{ is even function} \right]$$

$$= \frac{2b}{\sqrt{2\pi}} \cdot \frac{\sqrt{\pi}}{\sqrt{2}} \left[\because \int_0^\infty e^{-\frac{x^2}{2}} dx = \sqrt{\frac{\pi}{2}} \right]$$

$$= b$$

Mean, $\mu = b$

Variance of Normal Distribution

By definition, Variance $= E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-b)^2 e^{-1/2\left(\frac{x-b}{\sigma}\right)^2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^2 e^{-\frac{z^2}{2}} dz \left[\text{Putting } z = \frac{x-b}{\sigma} \text{ so that } dz = \frac{dx}{\sigma} \right]$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \cdot 2 \int_0^\infty z^2 \cdot e^{-z^2/2} dz \left[\because \text{Integrand is even function} \right]$$

$$= \frac{2\sigma^2}{\sqrt{2\pi}} \int_0^\infty 2t e^{-t} \frac{dt}{\sqrt{2t}} \left[\text{Putting } \frac{z^2}{2} = t \text{ so that } dz = \frac{dt}{\sqrt{2t}} \right]$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^\infty e^{-t} \sqrt{t}\, dt = \frac{2\sigma^2}{\sqrt{\pi}} \int_0^\infty e^{-t} t^{3/2-1} dt$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) = \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{1}{2} \cdot \Gamma\left(\frac{1}{2}\right)$$

$$= \frac{\cancel{2}\sigma^2}{\cancel{\sqrt{\pi}}} \cdot \frac{\cancel{\sqrt{\pi}}}{\cancel{2}}$$

$$= \sigma^2$$

Hence, variance $= \sigma^2$

Thus the standard deviation of the normal distribution is σ.

Median of Normal Distribution

If M is the median of the normal distribution, we have

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

i.e. $$\frac{1}{\sigma.\sqrt{2\pi}} \int_{-\infty}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{2}$$

$$\frac{1}{\sigma.\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx + \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{2} \quad \text{--- (1)}.$$

consider $$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

put $\frac{x-\mu}{\sigma} = z$, then $dx = \sigma\, dz$

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sigma.\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{z^2}{2}}.\sigma\, dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{z^2}{2}} dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{z^2}{2}} \text{(by symmetry)} = \frac{1}{\sqrt{2\pi}}\cdot\sqrt{\frac{\pi}{2}} = \frac{1}{2} \text{ --- (2)}$$

from (1) & (2), we hv

$$\frac{1}{2} + \frac{1}{\sigma.\sqrt{2\pi}} \int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{2}$$

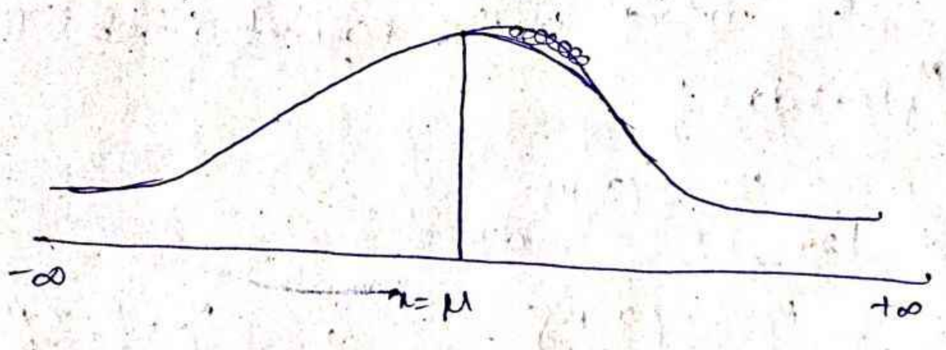
$$\frac{1}{\sigma.\sqrt{2\pi}} \int_{\mu}^{M} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} dx = 0$$

$$\int_{\mu}^{M} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 0 \Rightarrow \mu = M$$

[$\because$ if $\int_{a}^{b} f(x)\,dx = 0$ then $a=b$, where $f(x)>0$]

Hence for the normal distribution, Mean = Median

Chief - characteristics of Normal Distribution:



* The graph of normal distribution is in XY plane, is called Normal Curve.
* The curve is Bell shaped and symmetrical about the line $x = \mu$ and the two tails, one on the right and other on left, which are extends to ∞.
* The area under the normal curve represents the total population.
* The mean, median and mode of normal D are coinside, so the normal curve is known as unimodel.
* The probability that the normal varient X with mean μ and standard deviation σ lies between x_1 and x_2 is given by

$$P(x_1 \le X \le x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

where, the normal variat $z = \frac{x-\mu}{\sigma}$

* The area under the normal curve are distributed as follows

(i) Area of normal curve btwn $\mu \pm \sigma$ is 68.27%

i.e., $P(\mu - \sigma \le X \le \mu + \sigma) = 0.6827$

(ii) Area of normal curve btwn $\mu \pm 2\sigma$ is 95.43%

i.e., $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.9544$

(iii) Area of normal curve btwn $\mu \pm 3\sigma$ is 99.73%

i.e., $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.9973$

① For a normally distributed variant and mean 1 and standard deviation 3. find probability that

(i) $3.43 \leq X \leq 6.19$ (ii) $-1.43 \leq X \leq 6.19$

(iii) $X \geq 2.38$ (iv) $X \leq 1.28$

② If X is Normal variant with mean 30 & standard deviation 5. find probability that (i) $26 \leq X \leq 40$

(ii) $X \geq 45$ (iii) $X \leq 42$ (iv) $28 \leq X \leq 49$

③ If the masses of 300 students are normally distributed with mean 68 kg & standard deviation 3 kg, How many students have the masses

(i) Greater than 72 kg (ii) less than or equal to 64 kg

(iii) btwn 65 & 71 kg inclusive.

④ Suppose the weights of 800 students are normally distributed with mean 140 Pounds & S.D ~~to~~ 10 Pounds then find the no. of students whose weights are

(i) btwn 138 & 148 Pounds, (ii) More than 152 Pounds

(iii) fewer than 146 Pounds (iv) less than 137 Pounds

⑤ In a normal distribution 31% of item are under 45 & 8% are over 64, find the mean & variance of distribution.

⑥ In a normal distribution 7% of items are under 35 & 89% are under 63, Determine mean & variance of distribution.

① Given 'X' is a normal variant

mean, $\mu = 1$

Standard deviation (S.D), $\sigma = 3$

(i) To find $P(\underset{x_1}{3.43} \leq X \leq \underset{x_2}{6.19})$

when $x_1 = 3.43 \Rightarrow z_1 = \frac{x_1 - \mu}{\sigma}$

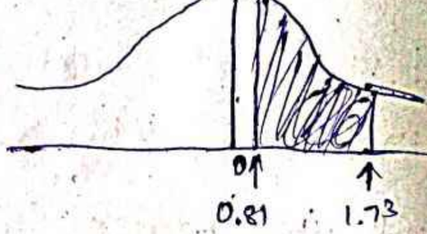
$= \frac{3.43 - 1}{3}$

$\boxed{z_1 = 0.81}$

when $x_2 = 6.19 \Rightarrow z_2 = \frac{x_2 - \mu}{\sigma}$

$= \frac{6.19 - 1}{3} = \frac{5.19}{3}$

$\boxed{z_2 = 1.73}$



$P(3.43 \leq X \leq 6.19) = P(0.81 \leq z \leq 1.73)$

$= A(0 \to 1.73) - A(0 \to 0.81)$

$= A(1.73) - A(0.81)$

$= 0.4582 - 0.2910$

$= 0.1672$

(ii) To find

$P(-1.43 \leq X \leq 6.19)$

when $x_1 = -1.43 \Rightarrow z_1 = \frac{x_1 - \mu}{\sigma}$

$= \frac{-1.43 - 1}{3}$

$\boxed{z_1 = -0.81}$

when $x_2 = 6.19 \Rightarrow z_2 = \frac{x_2 - \mu}{\sigma}$

$= \frac{6.19 - 1}{3}$

$\boxed{z_2 = 1.73}$

$P(-1.43 \le X \le 6.19) = P(-0.81 \le z \le 1.73)$

$= A(0 \to -0.81) + A(0 \to 1.73)$

$= A(-0.81) + A(1.73)$

Since area can't be negative

$= A(0.81) + A(1.73)$

$= 0.2910 + 0.4582$

$= 0.7492$

(iii) $X \ge 2.38$

when $x_1 = 2.38 \Rightarrow z_1 = \frac{x_1 - \mu}{\sigma} = \frac{2.38 - 1}{3} = \frac{1.38}{3}$

$z_1 = 0.46$

$P(X \ge 2.38) = P(z_1 \ge 0.46)$

$= 0.5 - A(0.46)$

$= 0.5 - 0.1772$

$= 0.3228$

(iv) $X \le 1.28$

when $x_1 = 1.28 \Rightarrow z_1 = \frac{x_1 - \mu}{\sigma} = \frac{1.28 - 1}{3} = \frac{0.28}{3}$

$z_1 = 0.0933$

$P(X \le 1.28) = P(z_1 \le 0.0933)$

$= 0.5 - 0.0359$

$= 0.4641$

2. Given 'x' is a normal varient

mean, $\mu = 30$ S.D, $\sigma = 5$

(i) To find $P(26 \le X \le 40)$

when $x_1 = 26 \Rightarrow z_1 = \frac{x_1 - \mu}{\sigma} = \frac{26 - 30}{5} = \frac{-4}{5}$

$\boxed{z_1 = -0.8}$

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

$\boxed{z_2 = 2}$

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

$= A(0 \to -0.8) + A(0 \to 2.00)$

Since area cant be negative

$= A(0 \to 0.8) + A(0 \to 2.00)$

$= 0.2881 + 0.4772$

$= 0.7653$

(ii) To find $X \ge 45$

$x = 45,\ z_1 = \frac{x - \mu}{\sigma} = \frac{45-30}{5} = 3$

$\boxed{z = 3}$

$P(X \ge 45) = P(z_1 \ge 3)$

$= 0.5 - A(0 \to 3)$

$= 0.5 - 0.4986$

$= 0.0013$

(iii) To find $X \le 42$

$x = 42,\ z = \frac{x - \mu}{\sigma} = \frac{42-30}{5} = \frac{8}{5} = 1.6$

$\boxed{z = 1.6}$

$P(X \le 42) = P(z \le 1.6)$

$= 0.5 - A(1.6)$

$= 0.5 - 0.4452$

$= 0.0548$

(iv) $28 \le X \le 49$

$x_1 = 28,\ z_1 = \frac{x_1 - \mu}{\sigma} = \frac{28-30}{5} = -\frac{2}{5}$

$\boxed{z_1 = -0.4}$

$x_2 = 49, z_2 = \frac{x_2 - \mu}{\sigma} = \frac{49-30}{5} = \frac{19}{5}$

$\boxed{z_2 \doteq 3.8}$

$P(28 \le X \le 49) = P(-0.4 \le z \le 3.8)$

$= A(0 \to -0.4) + A(0 \to 3.8)$

since area can't be negative

$= A(0.4) + A(3.8)$

$= 0.1554 + 0.4999$

$= 0.6553$

(3) Let x be masses of students.

$\mu = 68$ kgs $\sigma = 3$ kgs

(i) $X > 72$

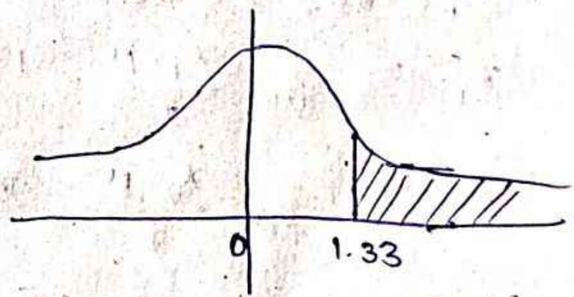
$x = 72, z = \frac{x-\mu}{\sigma} = \frac{72-68}{3} = 1.3333$

$P(X > 72) = P(z > 1.33)$

$= 0.5 - A(1.33)$

$\doteq 0.5 - 0.4082$

$= 0.0918$



number of students with more than 72 kgs $= 300 \times 0.0918 = 28$

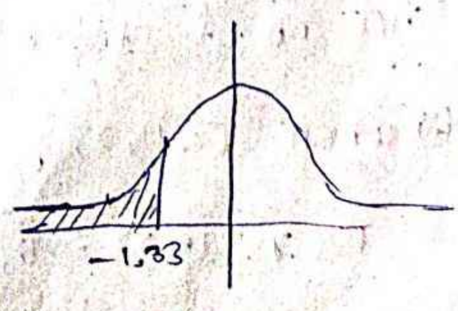
(ii) $X \le 64$

$x = 64, z = \frac{x-\mu}{\sigma} = \frac{64-68}{3} = -1.33$

$P(X \le 64) = P(z \le -1.33)$

$= 0.5 - A(1.33)$

$= 0.0918$



No. of students have masses less than or equal to 64 kg

$= 300 \times 0.0918 = 28$

(iii) $P(65 \le X \le 71)$

$x_1 = 65,\ z_1 = \frac{x_1 - \mu}{\sigma} = \frac{65-68}{3} = -1$

$x_2 = 71,\ z_2 = \frac{x_2 - \mu}{\sigma} = \frac{71-68}{3} = 1$

$P(65 \le X \le 71) = P(-1 \le z \le 1)$

$= A(0 \to -1) + A(0 \to 1)$

Area can't be negative

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

$= 0.3413 + 0.3413$

$= 0.6826$

Required no. of students $= 300 \times 0.6826 = 205$

(4) $\mu = 140$ pounds, $\sigma = 10$ pounds

(i) $P(138 \le x \le 148)$

$x_1 = 138,\ z_1 = \frac{x_1 - \mu}{\sigma} = \frac{138-140}{10} = -0.2$

$x_2 = 148,\ z_2 = \frac{x_2 - \mu}{\sigma} = \frac{148-140}{10} = 0.8$

$P(138 \le x \le 148) = P(-0.2 \le z \le 0.8)$

$= A(0 \to 0.2) + A(0 \to 0.8)$

$= A(0.2) + A(0.8)$

$= 0.0793 + 0.2881$

$= 0.3674$

Hence no. of students $= 0.3674 \times 800 = 294$

(ii) when $x = 152,\ z_1 = \frac{x - \mu}{\sigma} = \frac{152-140}{10} = 1.2$

$P(x > 152) = 0.5 - A(1.2)$

$= 0.5 - 0.3849$

$= 0.1151$

$\therefore$ No. of students $= 800 \times 0.1151 = 92$

(iii) fewer than 146 Pounds

$P(X<146)$

$x=146,\ z=\frac{x-\mu}{\sigma}=\frac{146-140}{10}=0.6$

$P(X<146)=P(z<0.6)$

$=0.5-A(0.6)$

$=0.5-0.2258$

$=0.2742$

(iv) less than 137 Pounds

$P(X<137)$

$x=137,\ z=\frac{x-\mu}{\sigma}=\frac{137-140}{10}=\frac{-3}{10}=-0.3$

$P(X<137)=P(z<-0.3)$

$=0.5-A(-0.3)$

$=0.5-A(0.3)$

$=0.5-0.1179$

$=0.3821$

Sample Distribution:

The totality of observations with which we are concerned, whether the number be B finite (or) infinite constitute is called Population.

* Population is the agrigate (or) totality of satistical data forming a subject of investigation.
* The no. of observations in the population is defined to be the size of population it may finite (or) infinite and is denoted by 'N'.
* The process of selection of a sample is known as sampling.
* The statistical constants like mean, median, mode, standard deviation, variance measures the population are called parameters and the measures obtained from the sample of population are called the statistic.

Sample Mean:

If $x_1, x_2, \ldots x_n$ represents a random sample of size 'n' then the sample mean defined by the statistic,

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

Sample Variance:

If $x_1, x_2, \ldots x_n$ represents a random sample of size 'n'. Then sample variance defined by the statistic σ^2 (or) s^2

$$\sigma^2 = \sum \frac{(x_i - \bar{x})^2}{n} \quad \text{(or)} \quad \sigma^2 = \sum \frac{(x_i - \bar{x})^2}{n-1}$$

Sample Standard deviation:

It is the positive square root of sample variance

* If N is the size of the population and n is the size of the sample, then

(i) The no. of samples using with replacement $= N^n$

(ii) The no. of samples using without replacement $= {}^N C_n$

→ The samples are classified into two ways, if the size of the sample $n \geq 30$ then the samples are said to be large samples.

→ If the size of the sample $n < 30$ then the sample is said to be small sample.

Central limit theorem:

If $\bar{x}$ be the mean of the sample of size n can be drawn from the population with mean(μ) and standard deviation (σ) then the standatised mean,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Sampling distribution of mean:

Infinite Population:

$$\text{Standard error} = \frac{\sigma}{\sqrt{n}}$$

Suppose the samples are drawn from an infinite population or sampling is done with replacement then

(i) Mean of the sampling distribution of means is given by $\mu_{\bar{x}} = \dfrac{\mu + \mu + \dots + \mu}{n}$

(ii) Variance of the sampling distribution of mean is given by $\sigma^2_{\bar{x}} = \dfrac{\text{Sum of squares of deviation from mean}}{n}$

i.e., $\sigma^2_{\bar{x}} = \dfrac{\sum(\mu_{x_i} - \mu_{\bar{x}})^2}{n}$ (or) $\sigma^2_{\bar{x}} = \dfrac{\sigma^2 + \sigma^2 + \dots + \sigma^2}{n} = \dfrac{\sigma^2}{n}$

(iii) Standard deviation of sampling distribution of mean is given by $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

Finite Population:

Suppose the samples are drawn from finite population are sampling without replacement then

(i) Mean of the sampling distribution of mean is given by $\mu_{\bar{x}} = \mu$

(ii) Variance of the sampling distribution of mean is given by $\sigma^2_{\bar{x}} = \dfrac{\sigma^2}{n}\left(\dfrac{N-n}{N-1}\right)$

(iii) Standard Deviation of sample distribution of mean is given by $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}\sqrt{\left(\dfrac{N-n}{N-1}\right)}$

Here $\dfrac{N-n}{N-1}$ is often called finite population correction factor.

① Find the finite populations correction factor $N = 200$, $n = 6$

Given, $N = 200$

$n = 6$

$$\frac{N-n}{N-1} = \frac{200-6}{200-1} = \frac{194}{199}$$

$$= 0.9748$$

② A Population consists of 5 no.s 2, 3, 6, 8 & 11 consider all possible samples of size 2 which can be drawn (i) With replacement (ii) Without replacement from this population. Find (i) Mean of the population (ii) S.D (iii) Mean of sampling distribution of means (iv) S.D of sampling distribution of means (standard error of population).

③ A population consists of 6 numbers 4, 8, 12, 16, 20, 24 consider all possible sample of size 2 which can be drawn with and without replacement from the population find (i) Mean of the population (ii) S.D (iii) Mean of Sampling distribution of means (iv) S.D of sampling distribution of means.

④ A population consists of 5 numbers 3, 6, 9, 15, 27 which can be drawn without replacement of size of sample is 3, then find

(i) Mean of the population (ii) S.D of the population
(iii) Mean of sampling distribution of means
(iv) S.D " " " " "

③ Using with replacement

Given, Population is 4, 8, 12, 16, 20, 24

size of the population, $N = 6$

size of the sample, $n = 2$

No. of samples using with replacement $= N^n = 6^2 = 36$

(i) Mean of the population $\mu = \frac{4+8+12+16+20+24}{6}$

$$\boxed{\mu = 14}$$

(ii) S.D of the population = ?

Variance of the population = σ^2

$$\sigma^2 = \frac{(4-14)^2+(8-14)^2+(12-14)^2+(16-14)^2+(20-14)^2+(24-14)^2}{6}$$

$$= \frac{(10)^2+(6)^2+(2)^2+(2)^2+(6)^2+(10)^2}{6}$$

$$= \frac{100+36+4+4+36+100}{6} = 46.6666$$

S.D of population $\sigma = \sqrt{46.6666} = 6.831$

The sampling distribution with $n=2$ is

(4, 4) (4, 8) (4, 12) (4, 16) (4, 20) (4, 24)

(8, 4) (8, 8) (8, 12) (8, 16) (8, 20) (8, 24)

(12, 4) (12, 8) (12, 12) (12, 16) (12, 20) (12, 24)

(16, 4) (16, 8) (16, 12) (16, 16) (16, 20) (16, 24)

(20, 4) (20, 8) (20, 12) (20, 16) (20, 20) (20, 24)

(24, 4) (24, 8) (24, 12) (24, 16) (24, 20) (24, 24)

Means of the samples are =

4	6	8	10	12	14
6	8	10	12	14	16
8	10	12	14	16	18
10	12	14	16	18	20
12	14	16	18	20	22
14	16	18	20	22	24

(iii) Mean of sampling distribution of means

$$\mu_{\bar{x}} = \frac{\begin{array}{c}4+6+8+10+12+14+6+8+10+12+14+16+8+10+12+14+16\\+18+10+12+14+16+18+20+12+14+16+18+20+22+14\\+16+18+20+22+24\end{array}}{}$$

$$\mu_x = \frac{504}{36}$$

$$\boxed{\mu_{\bar{x}} = 14}$$

(iv) Variance of sampling distribution of mean is

$$\sigma^2_{\bar{x}} = \frac{\Sigma(\mu_{xi} - \mu_x)^2}{n}$$

$$= \frac{\begin{array}{l}(4-14)^2 + (6-14)^2 + (8-14)^2 + (10-14)^2 + (12-14)^2 + (14-14)^2 + \\ (6-14)^2 + (8-14)^2 + (10-14)^2 + (12-14)^2 + (14-14)^2 + (16-14)^2 + \\ (8-14)^2 + (10-14)^2 + (12-14)^2 + (14-14)^2 + (16-14)^2 + (18-14)^2 + \\ (10-14)^2 + (12-14)^2 + (14-14)^2 + (16-14)^2 + (18-14)^2 + (20-14)^2 + \\ (12-14)^2 + (14-14)^2 + (16-14)^2 + (18-14)^2 + (20-14)^2 + (22-14)^2 + \\ (14-14)^2 + (16-14)^2 + (18-14)^2 + (20-14)^2 + (22-14)^2 + (24-14)^2\end{array}}{36}$$

$$= \frac{\begin{array}{l}100+64+36+16+4+64+36+16+4+4+36+16+4+4+16+16+ \\ 4+4+16+36+4+4+16+36+64+4+16+36+64+100\end{array}}{36}$$

$$= \frac{840}{36} = 23.3333$$

S.D of sampling distribution of mean (μ) is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4.8305}{\sqrt{36}} = 0.8051$$

Standard error of the population $= \frac{\sigma}{\sqrt{n}}$

$$= \frac{6.831}{\sqrt{2}}$$

$$= \frac{6.831}{1.414}$$

$$= 4.8303$$

Using without replacement:

(i) Mean of the population, $\mu = 14$

(ii) S.D of the population, $\sigma = 6.831$

(iii) $N = 6$, $n = 2$

The no.of samples using without replacement is,

$$^{N}C_{n} = {}^{6}C_{2} = 15$$

$$\cancel{(4,4)}\ \cancel{(4,8)}\ \cancel{(4,12)}\ \cancel{(4,16)}\ \cancel{(4,20)}\ \cancel{(4,24)}$$

$$(8,4)\ \cancel{(8,8)}\ \cancel{(8,12)}\ \cancel{(8,16)}\ \cancel{(8,20)}\ \cancel{(8,24)}$$

$$(12,4)\ (12,8)\ \cancel{(12,12)}\ \cancel{(12,16)}\ \cancel{(12,20)}\ \cancel{(12,24)}$$

$$(16,4)\ (16,8)\ (16,12)\ \cancel{(16,16)}\ \cancel{(16,20)}\ \cancel{(16,24)}$$

$$(20,4)\ (20,8)\ (20,12)\ (20,16)\ \cancel{(20,20)}\ \cancel{(20,24)}$$

$$(24,4)\ (24,8)\ (24,12)\ (24,16)\ (24,20)\ \cancel{(24,24)}$$

Mean of the samples are

$$= \frac{6+8+10+10+12+14+12+14+16+18+14+16+18+20+22}{15}$$

$$= 14$$

Mean of the sampling distribution of mean,

$$\mu_{\bar{x}} = \mu$$

$$\mu_{\bar{x}} = 14$$

(iv) variance of sampling distribution of mean is

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{46.6627}{2}\left(\frac{6-2}{6-1}\right)$$

(5) The mean height of the students in a college is 155 cm and S.D is 1.5 cm, what is the probability that the mean height of 36 students is less than 157 cm and more than 152 cm.

(6) Random sample of size 100 is taken from infinite population having the mean 76 with variance 256. What is the probability that $\bar{x}$ will be btwn 75 & 78.

(7) Random sample of size 64 is taken from a normal population with mean 51.4 and S.D is 6.8, what is the probability that the mean of the sample will

(i) exceeds 52.9 (ii) fall btwn 50.5 & 52.3

(iii) fewer than 50.6

Given, x is a normal variant

$\mu = 51.4 \quad S.D = 6.8 \quad n = 64$

The normalized variable, $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

(i) To find $P(\bar{x} \geq 52.9)$

when $\bar{x} = 52.9 \Rightarrow z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{52.9 - 51.4}{6.8/\sqrt{64}} = 1.76$

$P(\bar{x} \geq 52.9) = P(z \geq 1.76)$

Area of $(1.76) = 0.4608$

$= 0.5 - 0.4608$

$= 0.0392$

(ii) $P(50.5 \le \bar{x} \le 52.3)$

$$z_1 = \frac{50.5 - 51.4}{6.8/\sqrt{64}} = -1.06 \quad z_2 = \frac{52.3 - 51.4}{6.8/\sqrt{64}} = 1.06$$

$= (-1.06 \to 0) + (0 \to 1.06)$

$= 0.3554 + 0.3554$

$= 0.7108$

(iii) $P(\bar{x} \le 50.6)$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{50.6 - 51.4}{6.8/8} = -0.94$$

$P(\bar{x} \le 50.6) = P(z \le -0.94)$

Area of $(-0.94) = 0.3264$

$= 0.5 - 0.3264$

$= 0.1736$ //

(2) Sol: Given population is 2, 3, 6, 8 and 11

size of the population, $N = 5$

size of the sample, $n = 2$

No. of samples using with replacement $= N^n = 5^2 = 25$

<u>with replacement</u>

(i) Mean of population, $\mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$

(ii) S.D of the population

Variance of the population $= \sigma^2$

$$\sigma^2 = \frac{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2}{5} = 10.8$$

S.D, $\sigma = \sqrt{10.8} = 3.29$

The sampling distribution with $n=2$ is

(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)

Mean's of samples are =

2	2.5	4	5	6.5
2.5	3	4.5	5.5	7
4	4.5	6	7.0	8.5
5	5.5	7	8	9.5
6.5	7	8.5	9.5	11

(iii) Mean of the sampling distribution of means

$$\mu_{\bar{x}} = \frac{2+2.5+4+5+6.5+2.5+3+4.5+5.5+7+4+4.5+6+7.0+8.5+5+5.5+7+8+9.5+6.5+7+8.5+9.5+11}{25}$$

$$= \frac{150}{25} = 6 \quad \boxed{\mu_{\bar{x}} = 6}$$

(iv) variance of sampling distribution of mean is

$$\sigma^2_{\bar{x}} = \frac{\Sigma(\mu_{x_i} - \mu_{\bar{x}})^2}{n}$$

$$= \frac{\begin{gathered}(2-6)^2+(2.5-6)^2+(4-6)^2+(5-6)^2+(6.5-6)^2+(2.5-6)^2+\\(3-6)^2+(4.5-6)^2+(5.5-6)^2+(7-6)^2+(4-6)^2+(4.5-6)^2+(6-6)^2+\\(7-6)^2+(8.5-6)^2+(5-6)^2+(5.5-6)^2+(7-6)^2+(8-6)^2+(9.5-6)^2+\\(6.5-6)^2+(7-6)^2+(8.5-6)^2+(9.5-6)^2+(11-6)^2\end{gathered}}{25}$$

$$= \frac{135}{25} = 5.40.$$

S.D of sampling distribution of mean (μ) is

$\sigma_{\bar{x}} = \sqrt{5.40} = 2.32$

standard error of the population $= \frac{\sigma}{\sqrt{n}}$

$= \frac{3.29}{\sqrt{2}} = 2.3264$

S.D of sampling distribution of mean(μ) is, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$= \frac{2.32}{\sqrt{25}} = 0.464$

Using without replacement:

(i) mean, $\mu = 6$

(ii) S.D, $\sigma = 3.29$

(iii) $N = 5$, $n = 2$

The no. of samples using without replacement is

$^NC_n = {}^5C_2 = 10$

(2, 3) (2, 6) (2, 8) (2, 11)

(3, 6) (3, 8) (3, 11)

(6, 8) (6, 11)

(8, 11)

Mean's of samples are =

2.5 4 5 6.5

4.5 5.5 7

7 8.5

9.5

The mean of the samples are $= \frac{2.5+4+5+6.5+4.5+5.5+7+8.5+9.5}{10}$

$= 6$

$\mu_{\bar{x}} = \mu \Rightarrow \mu_{\bar{x}} = 6$

Variance of sampling distribution of mean is

$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{10.8}{2}\left(\frac{5-2}{5-1}\right)$

$= \frac{10.8}{2} \times \frac{3}{4} = 4.05$

(5) Sol:

Mean of the population,

μ = mean height of students of a college = 155 cm

S.D population = σ = 15 cms

n = sample size = 36

$\bar{x}$ = mean of sample = 157

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

$$= \frac{157-155}{15/\sqrt{36}} = 0.8$$

$$P(\bar{x} \le 157) = P(z < 0.8)$$
$$= 0.5 + P(0 \le z \le 0.8)$$
$$= 0.5 + 0.2881$$
$$= 0.7881$$

Thus the probability that the mean height of 36 students is less than 157 = 0.7881

(6) Sol: $n = 100$ $\mu = 76$ $\sigma^2 = 256$

$\sigma = 16$

n is large, the sample mean $\bar{x} \sim N(\mu, \sigma^2/\sqrt{n})$

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

$$\bar{x}_1 = 75,\ z_1 = \frac{\bar{x}_1-\mu}{\sigma/\sqrt{n}} = \frac{75-76}{16\sqrt{100}} = -0.625$$

$$\bar{x}_2 = 78,\ z_2 = \frac{\bar{x}_2-\mu}{\sigma/\sqrt{n}} = \frac{78-76}{16\sqrt{100}} = 1.25$$

$$P(75 \le \bar{x} \le 78) = P(z_1 \le Z \le z_2) = P(-0.625 \le Z \le 1.25)$$
$$= 0.2334 + 0.3944$$
$$= 0.6628$$

UNIT-4

Estimation: Quantities appearing in the distribution, such as p in the bionomial distribution, λ in Poison D μ, σ in Normal Distribution are called Parameters of the distribution.

→ The process or rule to determine an unknown population parameter is called an Estimator.

→ There are 2 kinds of estimations to determine the statistic of the population parameters namely,

(i) Point Estimation
(ii) Interval Estimation

Point Estimation:

→ If an estimate of a population parameter is given by a single value, the estimate is known as point estimate of the parameter and it will be denoted by $\hat{\theta}$.

→ A statistic or a point estimator $\hat{\theta}$ is said to be an unbiased estimator of the parameter θ if $E(\hat{\theta}) = \theta$.

Interval Estimation:

→ An interval estimate of a population parameter θ is an interval of the form $\hat{\theta}_L < \theta < \hat{\theta}_U$ where $\hat{\theta}_L$ & $\hat{\theta}_U$ depends on the value of static $\hat{\theta}$.

→ We shall able to determine $\hat{\theta}_L$ & $\hat{\theta}_U$ such that

$$P(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1 - \alpha \text{ for } 0 < \alpha < 1$$

→ The interval $\hat{\theta}_L < \theta < \hat{\theta}_U$ computed from the selected sample is called a $(1-\alpha)100\%$ confidence interval.

→ The error of the estimate is given by

$$E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

→ The number of samples in the estimation is given by

$$n = \left[\frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{t} \right]^2$$

→ A $(1-\alpha)100\%$ confidence intervals are feducial limits are given by

$$\bar{x} - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < x + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

→ The critical values of Z_α is given by for

$90\% \rightarrow Z_{\alpha/2} = 1.645$

$95\% \rightarrow Z_{\frac{\alpha}{2}} = 1.96$

$98\% \rightarrow Z_{\frac{\alpha}{2}} = 2.33$

$99\% \rightarrow Z_{\frac{\alpha}{2}} =$ ~~2.1645~~ 2.58

Bayesian ~~Estimation~~

Let μ_0 and σ_0 be the (μ) mean and S.D's of a subjective prior distribution.

→ Combining the prior possible values of μ and σ with the direct sample evidence, the posterior mean and S.D in Bayesian Estimation is approximated by the normal distribution with

* Posterior mean $(\mu_1) = \dfrac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$

* Posterior S.D $(\sigma_1) = \sqrt{\dfrac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}}$

* A $(1-\alpha)100\%$ Bayesian interval about the mean μ is given by

$$(1-\alpha)100\% \Rightarrow \mu_1 - Z_{\alpha/2} \cdot \sigma_1 < \mu < \mu_1 + Z_{\alpha/2} \cdot \sigma_1$$

① What is the maximum error one can expect to make the property 0.90 using the mean of random sample of size 64 to estimate mean of the population with valiance 2.56.

Given $n = 64$

Valiance$(\sigma^2) = 2.56$

S.D$(\sigma) = 1.6$

$90\% \Rightarrow Z_{\alpha/2} = 1.645$ $\boxed{E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}}$

$$E = 1.645 \times \frac{1.6}{\sqrt{64}} = 0.329$$

② A Random sample of size 100 has a S.D of 5. What can you say about the maximum number with 95% confidence.

③ Assuming that sigma $= 20$, How large a Random sample be taken to assert with probability 0.98 that the sample mean will not defer from the true mean by more than 3.0 points.

④ What is the size of the smallest sample required to estimate an unknown population with S.D of 0.25 within a maximum error of 0.06 of atleast 99% confident.

⑤ The mean and S.D's of the population are 11,795 and 14,054 respectively. If $n = 50$, construct 90%, 95%, 98%, 99% confidence intervals.

Sol: Mean of the population, $\mu = 11{,}795$

$S.D\ (\sigma) = 14{,}054$

$n = 50$

$90\% \rightarrow Z_{\alpha/2} = 1.645$

$95\% \rightarrow Z_{\alpha/2} = 1.96$

$98\% \rightarrow Z_{\alpha/2} = 2.33$

$99\% \rightarrow Z_{\alpha/2} = \cancel{2.1644}\ 2.58$

(i) 90% confidence interval is given by

$$\left[\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{50}},\ \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

$$\left[11{,}795 - (1.645)\frac{14054}{\sqrt{50}},\ 11{,}795 + (1.645)\frac{14{,}054}{\sqrt{50}}\right]$$

$$= (8525.5037,\ 15064.4962)$$

(ii) 95% confidence interval is given by

$$\left[11{,}795 - (1.96)\frac{14054}{\sqrt{50}},\ 11795 + (1.96)\frac{14{,}054}{\sqrt{50}}\right]$$

$$= (7899.43,\ 15690.57)$$

(iii) 98% confidence interval is given by

$$\left[11{,}795 - (2.33)\frac{14054}{\sqrt{50}},\ 11795 + (2.33)\frac{14054}{\sqrt{50}}\right]$$

$$= (7164.041,\ 16425.96)$$

(iv) 99% confidence interval is given by

$$\left[11795-(2.58)\frac{14054}{\sqrt{50}},\ 11795+(2.58)\frac{14054}{\sqrt{50}}\right]$$

(2) Given $n=100$

$S.D = 5$

$e = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$95\% = Z_{\alpha/2} = 1.96$

$\epsilon = Z_{\alpha/2} \cdot \frac{5}{\sqrt{100}} = 0.98$

(3) Given $\sigma = 20$

$n = ?$

$98\% \Rightarrow Z_{\alpha/2} = 2.33$

$\epsilon = 3$

$$n = \left[\frac{Z_{\alpha/2} \cdot \sigma}{\epsilon}\right]^2 = \left[\frac{2.33(20)}{3}\right]^2 = 241$$

(4) Given,

$S.D = 0.25$

$\epsilon = 0.06$

$99\% \Rightarrow Z_{\alpha/2} = 2.58$

$$n = \left[\frac{Z_{\alpha/2} \cdot \sigma}{\epsilon}\right]^2$$

$$n = \left[\frac{2.58 \cdot (0.25)}{0.06}\right]^2$$

$= 115.5625$

⑥ A professors feelings about the mean mark in the final examination in probability of a large group of students is expressed subjectively by normal distribution with $\mu_0 = 67.2$ and $\sigma_0 = 1.5$

(i) If the mean mark lies in the interval (65, 70) determine the prior probability of professors should assign to the mean mark.

(ii) Find the professors posterior mean and posterior S.D. If the examinations are conducted on a random sample of 40 students yielding mean 74.9 & S.D = 7.4

(iii) Determine the posterior probability which he will assigned the mean mark being in interval (65, 70)

(iv) Construct a 90%, 95%, 98%, 99% Bayesian intervals for μ.

(i) To find prior probability $P(65 \le x \le 70)$

when $x_1 = 65 \Rightarrow z_1 = \dfrac{x_1 - \mu_0}{\sigma_0}$

$$= \frac{65 - 67.2}{1.5} = -1.47$$

$x_2 = 70 \Rightarrow z_2 = \dfrac{70 - 67.2}{1.5} = 1.87$

$P(65 \le x \le 70) = P(-1.47 \le z \le 1.87)$

$= A(-1.47) + A(1.87)$

$= A(1.47) + A(1.87)$

$= 0.4292 + 0.4693$

$= 0.8985$

$\therefore$ Prior probability in the interval (65, 70) is 0.8985.

(ii) $\mu_1 = ?$ $\sigma_1 = ?$

$n = 40$

Sample mean$(\mu_{\bar{x}}) = 74.9$

$S.D = \sigma = 7.4$

$$\text{Posterior mean } \mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$$

$$= \frac{40(74.9)(1.5)^2 + 67.2(7.4)^2}{40(1.5)^2 + (7.4)^2}$$

$$= 71.99$$

$$\text{Posterior S.D } \sigma_1 = \sqrt{\frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}} = \sqrt{\frac{(1.5)^2(7.4)^2}{40(1.5)^2 + (7.4)^2}}$$

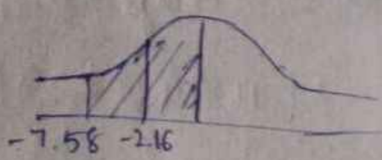
$$= 0.9225$$

(iii) To find the posterior probability $P(65 \le x \le 70)$

when $x_1 = 65 \Rightarrow z_1 = \dfrac{x_1 - \mu_1}{\sigma_1}$

$$= \frac{65 - 71.99}{0.9225}$$

$$= -7.58$$



$x_2 = 70 \Rightarrow z_2 = \dfrac{x_2 - \mu_1}{\sigma_1} = \dfrac{70 - 71.99}{0.9225}$

$$= -2.16$$

$\therefore P(65 \le x \le 70) = P(-7.59 \le z \le -2.16)$

$= A(-7.59) - A(-2.16)$

$= A(7.59) - A(2.16)$

$= 0.5 - 0.4846$

$= 0.0154$

$\therefore$ Posterior probability b/w (65 & 70) is 0.0154

(iv) We know that,

$Z_{\alpha/2}$ for 90% $= Z_{\alpha/2} = 1.645$

$Z_{\alpha/2}$ for 95% $= 1.96$

$Z_{\alpha/2}$ for 98% $= 2.33$

$Z_{\alpha/2}$ for 99% $= 2.58$

90% Bayesian interval about mean μ is given by

$(\mu_1 - Z_{\alpha/2} \cdot \sigma_1 , \mu_1 + Z_{\alpha/2} \cdot \sigma_1)$

$= [71.99 - 1.645(0.9225), 71.99 + 1.645(0.9225)]$

$= (70.47, 73.51)$

95% Bayesian interval

$= [71.99 - 1.96(0.9225), 71.99 + 1.96(0.9225)]$

$= (70.18, 73.80)$

98% Bayesian interval

$= [71.99 - 2.33(0.9225), 71.99 + 2.33(0.9225)]$

$= (69.84, 74.14)$

99% Bayesian interval

$= [71.99 - 2.58(0.9225), 71.99 + 2.58(0.9225)]$

$=$

Testing of Hypothesis

A statistical hypothesis is a quantative statement about a population parameter. Hypothesis is an assumption or a statement which is may or may not be true. It is a conclusion which is tentatively drawn as a logical basis.

The decision making procedure about the hypothesis is called hypothesis testing.

There are two types of hypothesis,

(i) Null hypothesis: The hypothesis formulated for the purpose of its rejection under the assumption that it is true, is called the null hypothesis and is denoted by H_0.

(ii) Alternative hypothesis: Any hypothesis which is complementary to the null hypothesis is called alternative hypothesis and is denoted by H_1.

Suppose to test the hypothesis that the population mean is equal to the hypothesized mean 60.

The possible alternative hypothesis can be stated as one of the following,

$H_1 : \mu \neq \mu_0$ (Two Tailed Test)

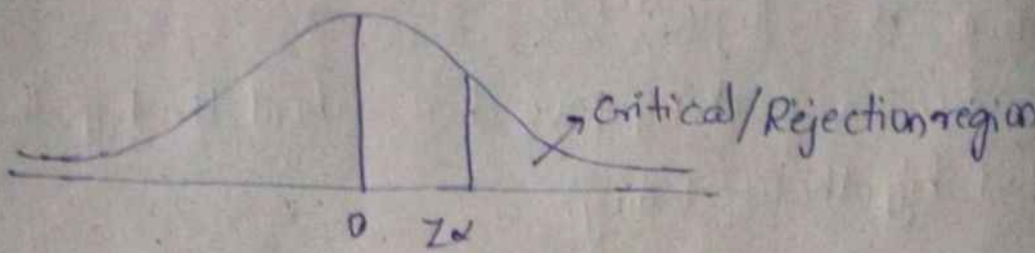
$H_1 : \mu > \mu_0$ (Right Tailed Test)

$H_1 : \mu < \mu_0$ (Left Tailed Test)

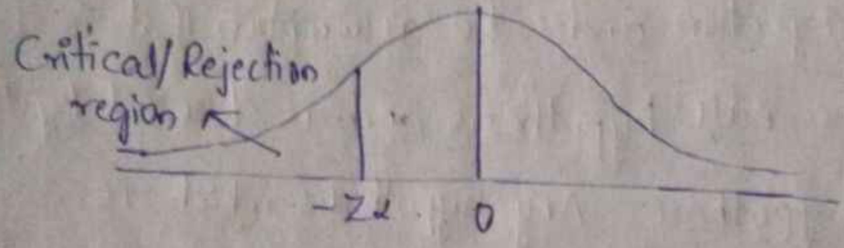
One tailed test:

A test of any statistical hypothesis where the alternative hypothesis is one sided as, $\mu > \mu_0$ or $\mu < \mu_0$ is called one sided test.

→ In right tailed test, the rejection region or the critical region lies entirely on the right side under the normal curve.



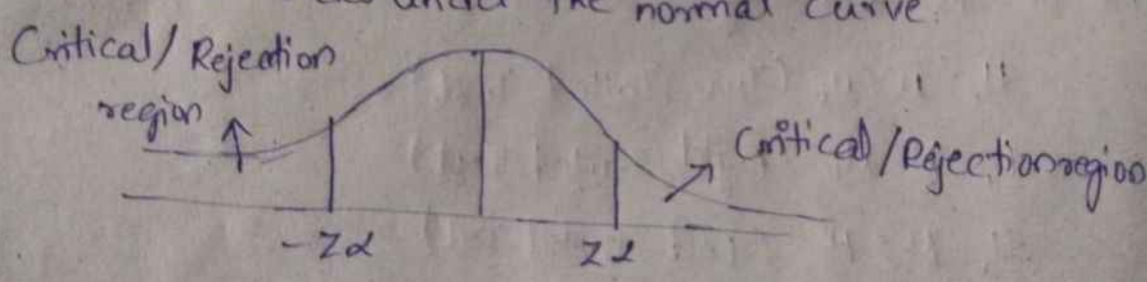
→ In left tailed test, the rejection region or the critical region lies entirely on the left side under the normal curve



Two tailed test:

A test of any statistical hypothesis where the alternative hypothesis is two sided such as $\mu \neq \mu_0$ is called two sided test.

In two tailed test, there are two rejection regions, one on each sides under the normal curve.



Critical region or Rejection region:

A region corresponding to a statistic which leads to rejection of the null hypothesis H_0 is called a Critical region or rejection region.

Level of Significance (α):

The level of significance is the maximum probability of a rejecting a null hypothesis when it is true and is denoted by 'α'.

The level of significances are 1% ($\alpha = 0.01$), 5% ($\alpha = 0.05$) and 10% ($\alpha = 0.1$).

When the level of significance is not mentioned, it can be considered as 5% ($\alpha = 0.05$) generally.

The procedure for testing of hypothesis:

(i) Null hypothesis (H_0): We set up the null hypothesis as the positive statement/condition from the given data as H_0.

(ii) Alternative hypothesis (H_1): We set up the alternative hypothesis as complementary statement or condition to null hypothesis as H_1.

Here, decide whether, we use 2 tailed or one (right/left) tailed test.

(iii) Level of significance (α):

Here, select the appropriate level of significance α

(iv) Test statistic:

Compute the test statistic using the appropriate method and formula (z or t or F or χ^2)

(v) Conclusion:

Compare the calculated value (C.V) with tabulated value (T.V)

If C.V < T.V → we accept the H_0

If C.V > T.V → we reject the H_0

Errors in sampling:

	Accepted	Rejected
H_0 is true	Correct decision	TYPE-I Error
H_0 is false	TYPE-II Error	Correct decision

→ There are 2 types of samples.

(i) Large Samples

(ii) Small Samples

To test the significance of large samples, we have large sample test (Z). To find the significance of small samples, we have

(i) Students t - test

(ii) Snedecor's F - test

(iii) Chi - Square test (χ^2)

(1) Test of significance for single mean:

When $\bar{x}$ is the sample mean, μ is the Population mean, σ is the standard deviation and n is the size of the sample, then the test statistic for significance of single mean is given by

$$\boxed{z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}$$

The confidence intervals are given by,

$$\left[\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Critical values of Z_α are given by.

$(Z_{\alpha/2})$	level of significance (α)		
	1% (0.01)	5% (0.05)	10% (0.1)
TTT	2.58	1.96	1.645
RTT	2.33	1.645	1.28
LTT	-2.33	-1.645	-1.28

① ~~A~~ sample of 64 students have a mean ~~wei~~ weight of ~~60~~ 70 kgs can this be regaeded as a sample from ~~the~~ population with mean weight 56 kgs & S.D(σ) 25 kgs.

Given, Sample size $n = 64$, is a large sample

Sample mean, $\bar{x} = 70$ kgs

population mean, $\mu = 56$ kgs

standaed deviation, $\sigma = 25$ kgs

Since sample mean $\bar{x}$ and population mean μ are available so we use, test of significance for single mean with

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{70 - 56}{25/\sqrt{64}}$$

Null hypothesis (H_0), The sample be regaeded from a population with mean weight 56 kgs

$$\mu = 56 \text{ kgs}$$

Alteenative hypothesis (H_1), The sample cannot be regarded from a population with mean weight 56kgs

$$\mu \neq 56 \text{ kgs}$$

which is a two tailed test (TTT)

Level of Significance:

The standard level of significance, $\alpha = 0.05$

Test Statistic is given by

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{70-56}{25/\sqrt{64}} = 4.48$$

$$z_{cv} = 4.48$$

$$z_{TV} = 1.96$$

$z_{cv} > z_{TV}$, z_{cv} is rejected

So, alternative hypothesis is accepted.

(2) An ambulance service claims that it takes less than 10 minutes in emergency calls. A sample of 36 calls has a mean time of 11 minutes with ~~SD~~ variance of 16 minutes. Test the claim at 5% level

$n = 36$ (large sample i.e., ≥ 30)

Sample mean, $\bar{x} = 11$ min

Variance $= \sigma^2 = 16$

Standard Deviation $= \sigma = \sqrt{16} = 4$ min

$\alpha = 5\%$

population mean, $\mu = 10$ min

Since sample and population mean's are available so we use test of significance for single mean with

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Null hypothesis (H_0): An ambulance service claims takes 10 minutes to reach its destination in emergency calls. $\mu = 10$ min.

Alternative Hypothesis (H_1): An ~~annual~~ ambulance service ~~claims~~ takes less than 10 min to reach its destination in emergency call.

$$\mu < 10 \text{ min.}$$

which is LTT (Left Tailed Test)

Level of significance:

Given level of significance is $\alpha = 5\% = 0.05$

Test statistic is given by

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{11 - 10}{4/\sqrt{36}} = \frac{3}{2} = 1.5$$

$$|z| = 1.5$$

Conclusion: $z_{cv} = 1.5$

$$z_{TV} = z_{0.05, LTT} = -1.645$$

Since $z_{cv} > z_{TV}$, we reject the null hypothesis (H_0)

∴ We accept Alternative Hypothesis (H_1)

An ambulance service takes less than 10 min to reach its destination in emergency call.

$$\mu < 10 \text{ min}$$

(3) According to the norms established for a mechanical ~~amplitude~~ aptitude test, the persons who are 18 years old have an average height of 73.2 with S.D of 8.6 if 40 persons selected randomly of that age averaged 76.7, test the hypothesis at $\mu = 73.2$ against the alternative hypothesis $\mu > 73.2$ at 0.01 level of significance

1) A sample of 900 members has a mean of 3.4 cm's with S.D of 2.61 centimeters (cm's). Is the sample from a large population of mean 3.25 cm's, if the population is normal and find 95% fiducial limits of true mean.

$$n = 900 \text{ (large sample)}$$

sample mean, $\bar{x} = 3.4$ cm's

$S.D, \sigma = 2.61$ cm's

population mean, $\mu = 3.25$ cm's

$$\alpha = 95\%$$

Since sample & population mean's are available so we use test of significance for single mean.

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Null Hypothesis (H_0): The sample is from large population of mean 3.25 cm's

$$\mu = 3.25$$

Alternative Hypothesis (H_1): The sample is not from large population with mean 3.25

$$\mu \neq 3.25$$

which is TTT (Two Tailed Test)

Level of significance:

Given level of significance is $\alpha = 0.05$

Test statistic is given by

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{3.4 - 3.25}{2.61 / \sqrt{900}} = 1.7241$$

Conclusion: $Z_{CV} = 1.741$

$$Z_{TV} = Z_{0.05,\,TTT} = 1.96$$

Since $Z_{CV} < Z_{TV}$, ~~and~~ so we accept null hypothesis (H_0)

The sample is from large population with mean $= 3.25$

95% fiducial / confidential limits are given by

$$\cancel{z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}} \quad \cancel{\frac{\bar{x}+\mu}{\sigma/\sqrt{n}}} \quad \left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

$$\left[3.4 - (1.96)\frac{2.61}{\sqrt{900}},\ 3.4 + (1.96)\frac{2.61}{\sqrt{900}}\right]$$

$$(3.2295,\ 3.5705)$$

~~Test of significance for difference of means:~~

(3) Sol: $n = 4$

Sample mean, $\bar{x} = 76.7$

Population mean, $\mu = 73.2$

S.D, $\sigma = 8.6$

$\alpha = 0.01$

Since sample and population mean's are available so we use test of significance for single mean.

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Null Hypothesis (H_0):

$$\mu = 73.2$$

Alternative hypothesis: $\mu > 73.2$

which is RTT (Right-tailed test)

Level of significance: $\alpha = 0.01$

The test statistic is,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{76.7 - 73.2}{8.6/\sqrt{4}}$$

$$z = 0.814$$

$$z_{cv} = 0.814$$

$$z_{TV} = 2.33$$

$z_{cv} < z_{TV}$, so null hypothesis H_0 is accepted

That is, $\bar{x}$ and μ donot differ significantly.

② Test of Significance for difference of mean's: (large sample)

The test statistic for difference of mean's is given by,

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

where, n_1, n_2 are size's of the samples

$\bar{x}, \bar{y}$ mean's of the sample

σ_1, σ_2 are S.D's

① The mean's of 2 large samples of sizes 1000 & 2000 members are ~~60~~ 67.5 inches and 68 inches respectively. Can the sample be regarded as drawn from the same population of S.D 2.5 inches.

Given, $n_1 = 1000$ (large samples)

$n_2 = 2000$

Sample mean, $\bar{x} = 67.5$ inches

$\bar{y} = 68$ inches

S.D, $\sigma = 2.5$ inches

Since 2 samples and their mean's are available so we use test of ~~sing~~ significance of difference of means

$$z = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

Null Hypothesis (H_0):

The sample are regarded from the same population.

$\bar{x} = \bar{y}$

Alternative Hypothesis (H_1)

The sample are not regarded from the same population

$\bar{x} \neq \bar{y}$

which is TTT (Two Tailed Test)

Level of significance: $\alpha = 0.05$

Test significance is given by

$$z = \frac{67.5-68}{\sqrt{\frac{2.5^2}{1000}+\frac{2.5^2}{2000}}}$$

$z = -5.1639$

$|z| = 5.1639$

Conclusion: $z_{cv} = 5.1639$

$z_{TV} = z_{0.05, TTT} = 1.96$

Since $z_{cv} > z_{TV}$, we reject z_{cv} & accept z_{TV}

The sample are not regarded from the same population.

② Q : A researcher wants to know the intelligence of the students in a school, he selected 2 groups of students ~~in a school~~ In first group there are 150 students having mean IQ of 75 & SD of 15. whereas in 2nd group there are 250 students having mean IQ of 70 & SD of 20.

Given $n_1 = 150$, $n_2 = 250$

Sample mean's, $\bar{x} = 75$, $\bar{y} = 70$

S.D's, $\sigma_1 = 15$, $\sigma_2 = 20$

Since we have 2 samples and their mean's and S.D's are available so we use test of significance of difference mean's

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Null Hypothesis (H_0): The groups have been came from the same population.

i.e., $\mu_1 = \mu_2$

Alternative Hypothesis (H_1): The groups doesn't ~~have~~ came from the same population

$\mu_1 \neq \mu_2$

Level of significance : $\alpha = 0.05$

Test statistic

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{75 - 70}{\sqrt{\frac{15^2}{150} + \frac{20^2}{250}}}$$

$$z = 2.7116$$

Conclusion: $Z_{cv} = 2.7116$

$$Z_{TV} = Z_{0.05, TTT} = 1.96$$

$Z_{CV} > Z_{TV}$, so we reject null hypothesis (H_0)

The groups doesnot come from the same population.

③ ~~A~~ sample of heights of 6400 english man has a mean of 67.85 inches with S.D of 2.56 inches. While a sample of heights of 1600 Austrians has a mean of 68.55 inches with S.D of 2.52 inches. Do the data indicates that the Austrians are on the Average then english man at 1%

Eng

$n_1 = 6400 \quad n_2 = 1600$

$\bar{x} = 67.85 \quad \bar{y} = 68.55$

$\sigma_1 = 2.56 \quad \sigma_2 = 2.52$

Since 2 samples and their means are available so we use significance for differences of mean's with

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Null hypothesis (H_0): The heights of austrians on the average equal of the english man

$$\bar{x} = \bar{y}$$

Alternative hypothesis (H_1): The heights of englishman are shorter than austrian man

$$\bar{x} < \bar{y} \quad (LTT)$$

Level of significance:

$$\alpha = 0.01$$

Test statistic:

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$z = \frac{67.85 - 68.55}{\sqrt{\frac{(2.56)^2}{6400} + \frac{(2.52)^2}{1600}}}$$

$$z = -9.906$$

$$|z| = 9.906$$

Conclusion:

$$Z_{CV} = 9.906$$

$$Z_{TV} = Z_{0.01, LTT} = -2.33$$

Since $Z_{CV} > Z_{TV}$, so we reject Null hypothesis

∴ we acept alternative hypothesis

The heights of english man shorter than Austrian man.

$$\bar{x} < \bar{y}$$

(3) Test of Significance for Single Proportion

→ The test statistic is given by $z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}}$

Here, P = population proportion (If not given consider 50%)

sample proportion, $p = \dfrac{x}{n}$ x → no. of outcomes, n → Total no. of samples

$$Q = 1 - P$$

→ confidencial limits are given by

$$\left(P - z_{\alpha/2} \sqrt{\frac{PQ}{n}},\ P + z_{\alpha/2} \sqrt{\frac{PQ}{n}}\right)$$

1. Q: A manufaturer claimed that at least 95% of the equipment which he supplied to a factor are conformed to the specifications. An examination of sample of 200 pieces of equipments revealed that 18 are faulty. The test is claimed at 5% level.

$$n = 200$$

The no. of pieces conformed to the specification, $x = 200 - 18$
$x = 182$

Sample proportion, $p = \dfrac{x}{n} = \dfrac{182}{200} = 0.91$

Population proportion $= P = 95\% = 0.95$

$$\alpha = 5\% = 0.05$$

$$Q = 1 - P$$

$$Q = 0.05$$

Null hypothesis (H_0): 95% of the equipments are upto the specifications

$$P = 95\%$$

Alternative hypothesis (H_1):

Atleast 95% equipments are conformed to specifications

$$P < 95\%$$

which is LTT

Level of significance: $\alpha = 0.05$

Test statistic:

$$|z| = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.91 - 0.95}{\sqrt{\frac{0.05 \times 0.95}{200}}}$$

$$= \frac{-0.04}{0.0154} = 2.59$$

Conclusion: $Z_{CV} = 2.59$

$$Z_{TV} = Z_{0.05\,LTT} = -1.645$$

Since $Z_{CV} > Z_{TV}$ we reject H_0 & accept H_1,

Atleast 95% of equipment are conformed to specifications.

Confidencial intervals are

$$\left(p - z_{\alpha/2} \cdot \sqrt{\frac{PQ}{n}},\ p + z_{\alpha/2} \sqrt{\frac{PQ}{n}}\right)$$

$$= (0.95 - (-1.645)(0.0154),\ 0.95 + (-1.645)(0.0154))$$

$$=$$

(2) 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% of level of significance?

$n = 1000$ No. of rice eaters, $x = 540$

p = Sample porportion of rice eaters $= \frac{540}{1000} = 0.54$

P = Population proportion of rice eaters $= \frac{1}{2} = 0.5$

$Q = 1 - P = 0.5$

Null hypothesis (H_0):

Both rice and wheat are equally popular in the state.

$P = 0.5$

Alternative hypothesis (H_1):

Both rice and wheat are not equally popular in state.

$P \neq 0.5$

which is TTT

Level of significance: $\alpha = 0.01$

Test statistic:

$$|z| = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$= \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{1000}}} = 2.532$$

$$z = 2.532$$

Conclusion: $z_{cv} = 2.532$

$z_{TV} = z_{0.01, TTT} = 2.58$

$z_{cv} < z_{TV}$, we accept H_0 and reject H_1

Both rice and wheat are equally popular in state

(3) 30 people were attacked by the disease and 18 only survived. Will you reject the hypothesis that the survival rate if attacked by this disease is 85% in favour of the hypothesis that is more at 5% level.

$n = 30$

x = People survived = 18

p = sample proportion = $\frac{18}{30} = 0.6$

P, population proportion, P = 85%

$P = 0.85$

$Q = 1 - P = 1 - 0.85 = 0.15$

$\alpha = 5\% = 0.05$

Null hypothesis (H_0):

The proportion of persons survived after attack by a disease

$P = 0.85$

Alternative hypothesis (H_1)

The proportion of person does not survive after attack by disease.

$P > 0.85$

which is RTT

Level of significance: $\alpha = 0.05$

Test statistic:

$$|z| = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.6 - 0.85}{\sqrt{\frac{0.85 \times 0.05}{30}}}$$

$|z| = 6.6421$

Conclusion: $Z_{cv} = 6.6421$

$Z_{TV} = Z_{0.05, RTT} = 1.645$

$Z_{cv} > Z_{TV}$, so we reject H_0 & accept H_1

(4) Test of significance for difference of proportions:

Test statistic for difference of proportions is given by,

$$z = \frac{P_1 - P_2}{\sqrt{Pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

for population proportion

where n_1, n_2 are the large samples x_1, x_2 are the outcomes of the experiments.

$$z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

Sample proportion, $P_1 = \frac{x_1}{n_1}$, $P_2 = \frac{x_2}{n_2}$

probability of success, $P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$

$$\boxed{q = 1 - P}$$

(1) Two samples of 400 men & 600 women were asked whether they would like to have fly-over near their residence, 200 men & 325 women were in favour of the proposal. test the hypothesis that the proportions of men and women in favour of the proposal are same at 5% level

(2) In 2 large populations there are 30% & 25% respectively of fair haired people, is this difference likely to be hidden in the samples of 1200 & 900 respectively from the 2 populations.

(3) In a city A, 20% of random samples of 900 school boys has a certain slight physical defect and in another city B, 18.5% of random samples of 1600 school boys has same defect. Is the difference b/w the proportions is significant at 0.05% level?

3. Sol: $n_1 = 900$, $n_2 = 1600$

no. of students in city A with physical defect

$$x_1 = 900 \times 20\%$$

$$= \frac{900 \times 20}{100}$$

$$x_1 = 180$$

no. of students with physical defect in city B

$$x_2 = 1600 \times 18.5\%$$

$$x_2 = 296$$

here, $P_1 = \frac{x_1}{n_1} = \frac{180}{900} = 0.2$

$$P_2 = \frac{x_2}{n_2} = \frac{296}{1600} = 0.185$$

Since 2 samples and their proportions are available, we use test for significance for different proportions using

$$z = \frac{P_1 - P_2}{\sqrt{Pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$

$$= \frac{(900)(0.2) + (1600)(0.185)}{900 + 1600}$$

$$\boxed{P = 0.1904}$$

$$q = 1 - P = 1 - 0.1904$$

$$\boxed{q = 0.8096}$$

$z =$

Null hypothesis (H_0): The difference btwn the proportions is significant

$$P_1 = P_2$$

Alternative hypothesis (H_1): The difference btwn the proportions is not significant

$$P_1 \neq P_2$$

which is TTT (Two Tailed Test)

Level of significance : $\alpha = 0.05$

Test significance is given by

$$z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{0.2 - 0.185}{\sqrt{(0.1904)(0.8096)\left(\frac{1}{900} + \frac{1}{1600}\right)}}$$

$$= \frac{0.015}{\sqrt{(0.1541)\left(\frac{1}{900} + \frac{1}{1600}\right)}}$$

$$z = 0.9169$$

$$z_{CV} = 0.9169$$

$$z_{TV} = 1.96$$

Thus, $z_{CV} < z_{TV}$, $\cancel{H_0}$ z_{CV} is accepted

The difference btwn the propotions is significant.

(1) Sol:

$n_1 = 400 \quad n_2 = 600$

$x_1 = 200 \quad x_2 = 325$

$$P_1 = \frac{x_1}{n_1} = \frac{200}{400} = 0.5$$

$$P_2 = \frac{x_2}{n_2} = \frac{325}{600} = 0.5417$$

Since p_1, p_2 and n_1, n_2 are available we use significance of difference of proportions using

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$= \frac{400(0.5) + 600(0.5417)}{400 + 600}$$

$$p = 0.5250$$

$$q = 1 - p = 0.4750$$

Null hypothesis (H_0):

The proportions of men and women in favour of the proposal are same

$$p_1 = p_2$$

Alternative hypothesis (H_1)

$$p_1 \neq p_2$$

The proportions of men & women in favour of the proposal are not same

which is TTT

level of significance:

$$\alpha = 5\% = 0.05$$

Test of significance:

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$z = \frac{0.5 - 0.5417}{\sqrt{(0.5250)(0.4750)\left(\frac{1}{400} + \frac{1}{600}\right)}}$$

$z = -0.8714$

$|z| = 0.8714$

$z_{cv} = 0.8714 \quad z_{TV} = 1.96$

$z_{cv} < z_{TV}$, so z_{cv} is accepted

The proportions of men & women are in favour of proposal are same.

(2) sol: $n_1 = 1200$, $n_2 = 900$

The population proportion of 1200 people, $P_1 = 30\% = 0.3$

the population proportion of 900 people, $P_2 = 25\% = 0.25$

$Q_1 = 0.7 \quad Q_2 = 0.75 \ (1 - 0.25)$

Since population proportions P_1, P_2, Q_1, Q_2 and n_1, n_2 are available so we use significance difference b/w population proportions using $z = \dfrac{P_1 - P_2}{\sqrt{\dfrac{P_1Q_1}{n_1} + \dfrac{P_2Q_2}{n_2}}}$

Null hypothesis(H_0): The difference is likely to be hidden in the samples

$P_1 = P_2$

Alternative hypothesis (H_1): the difference is not likely to be hidden in the samples

$P_1 \neq P_2$.

which is TTT

Level of significance: standard level of significance is

$\alpha = 0.05\%$

Test of significance:

$$z = \frac{P_1 - P_2}{\sqrt{\dfrac{P_1Q_1}{n_1} + \dfrac{P_2Q_2}{n_2}}} = \frac{0.3 - 0.25}{\sqrt{\dfrac{0.3 \times 0.7}{1200} + \dfrac{0.25 \times 0.75}{900}}} = 2.5537$$

$z_{cv} = 2.5537 \quad z_{TV} = 1.96 \Rightarrow z_{cv} > z_{TV}$, so z_{cv} is rejected

Test of significance for small samples:

Students t-test / distribution

① Test of significance for single mean - small samples

Test statistic is given by, $t = \frac{\bar{x} - \mu}{S/\sqrt{n-1}}$ (when S.D is available)

when μ is available, $t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$ (when S.D is not available)

where $S^2 = \sum \frac{(x_i - \bar{x})^2}{n-1}$

Degrees of freedom $\nu = n-1$

Confidence interval is given by,

$$\left(\bar{x} - \frac{t_\alpha}{2} \cdot \frac{s}{\sqrt{n-1}}, \bar{x} + \frac{t_\alpha}{2} \cdot \frac{s}{\sqrt{n-1}}\right)$$

① A sample of 26 bulbs gives a mean life of 990 hours with S.D of 20 hours. The manufacturer claims that the mean life of bulbs is 1000 hours. Is the sample not up to the standards. Also construct the confidence intervals.

$n = 26$ (small sample < 30)

Sample mean, $\bar{x} = 990$

S.D, $s = 20$

population mean, $\mu = 1000$

Since sample and population mean's are available so we use significance of single mean using

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n-1}}$$

Degrees of freedom, $\nu = (n-1)$

$\nu = 20 - 1 = 19$

Null hypothesis: The sample is up to the standards.

$\mu = 1000$

Alternative hypothesis: The sample is not upto the standards.

$\mu < 1000$

which is LTT (left Tailed Test)

Level of significance: standard level of significance, $\alpha = 0.05$

Test statistic:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{990 - 1000}{20/\sqrt{26-1}} = \frac{-10}{20/5} = \frac{-10}{4}$$

$$= -2.5$$

$$|t| = 2.5$$

$t_{cv} = 2.5$

$t_{TV} = t_{\nu,\alpha} = t_{25, 0.05} = 1.708$

since $t_{cv} > t_{TV}$, so we reject null hypothesis.

The sample is not upto the standards

$\mu < 1000$

Fiducial / Confidential limits:

$$\left(\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n-1}}, \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n-1}}\right)$$

$$\left(990 - \overset{1.708}{(t_{0.025})} \cdot \frac{20}{\sqrt{26-1}}, 990 + 1.708 \cdot \frac{20}{\sqrt{26-1}}\right)$$

$$\left(990 - \overset{1.708}{t_{0.05}} \cdot \frac{20}{5}, 990 + 1.708 \cdot \frac{\overset{4}{\cancel{20}}}{\cancel{5}}\right)$$

(983.168 , 996.832)

(5) A mechanist is making engine parts with axial diameter 0.700 inches, a random sample of 10 parts shows the mean diameter of 0.742 inches with S.D of 0.040 inches. Compute the statistic you would use to test whether the work is meeting the specifications at 1% level.

$n = 10$ (small sample i.e., < 30)

Sample mean, $\bar{x} = 0.742$ inches

S.D, $s = 0.040$ inches

Population mean, $\mu = 0.700$ inches

Since sample and population mean's are available so we use significance of single mean using

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

Degrees of freedom, $\nu = n-1 = 10-1 = 9$

Null hypothesis (H_0)

The product is confirming to specification.

$$\mu = 0.700$$

Alternative hypothesis (H_1)

The product is not confirming to specification

$$\mu \neq 0.700$$

which is TTT

Level of Significance:

$$\alpha = 0.01$$

Test statistic:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{0.742 - 0.700}{0.040/\sqrt{10-1}} = 3.15$$

Conclusion: $t_{cv} = 3.15$

$$t_{tv} = t_{\nu,\alpha} = t_{9,0.01} = 2.821$$

$t_{cv} > t_{tv}$, so H_0 is rejected

The product is not meeting the specification.

(3) A random sample of 10 boys ait the following IQ's value 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do this data supports the assumption of population mean IQ of 100. find the reasonable range in which most of the mean IQ values of samples of 10 boys lie.

Given that, $n = 10$ (small sample - $n < 30$)

Population mean IQ, $\mu = 100$

IQ of 10 boys are

$$\bar{x} = \frac{70+120+110+101+88+83+95+98+107+100}{10}$$

Sample mean IQ, $\bar{x} = 97.2$

Since sample and population mean's are available so we use test of significance for single mean

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Degree of freedom $(\nu) = n-1 = 10-1 = 9$

x_i	$(x_i - \bar{x})$ $\bar{x} = 97.2$	$(x_i - \bar{x})^2$
70	-27.2	739.84
120	22.8	519.84

110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
		1833.6

$$S^2 = \frac{\Sigma(x_i - \bar{x})}{n-1} = \frac{1833.6}{9} = 203.733$$

$$S = \sqrt{203.733} = 14.2735$$

Null hypothesis: The data supports the assumption of population mean IQ of 100

$$\mu = 100$$

Alternative hypothesis: The data does not supports the assumption of population mean of IQ 100.

$$\mu \neq 100$$

which is TTT

Level of significance: $\alpha = 0.05$ (standard)

Test statistic:

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

$$= \frac{97.2 - 100}{14.27/\sqrt{10}} = -0.6204$$

$$|t| = 0.6204$$

Conclusion: $t_{cv} = 0.6204$

$t_{TV} = t_{\upsilon, \alpha/2}$

$= t_{9, 0.025}$

$t_{TV} = 2.262$

Since $t_{cv} < t_{TV}$ so we accept Null hypothesis.

The data supports the assumption of population mean IQ of 100

$\mu = 100$

→ The resonable range is given by

$$\left[\bar{x} - t_{\alpha/2} \cdot S/\sqrt{n}, \bar{x} + t_{\alpha/2} S/\sqrt{n} \right]$$

$$\left[97.2 - 2.262 \cdot \frac{14.2735}{\sqrt{10}} + 97.2 + 2.262 \frac{14.2735}{\sqrt{10}} \right]$$

$[86.99, 107.4074]$

(4) The life time of electric bulbs for a random sample of 10 from large consenment given the following data

Items	1	2	3	4	5	6	7	8	9	10
Life in 1000 hrs	4.2	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 average life time of bulbs is 4000 hrs.

Given $n = 10$ (small samples)

Null hypothesis (H_0): The average life time of bulbs is 4000 hrs

$\mu = 4000$ hrs

Alternative hypothesis (H_1): The average life time of bulbs is not equal to 4000 hrs

$\mu \neq 4000$ hrs.

which is TTT

Level of significance: $\alpha = 0.05$

Test of statistic:

$n = 10$

$\mu = 4000$

sample mean, $\bar{x} = \dfrac{1.2+4.6+3.9+4.1+5.2+3.8+3.9+4.3+4.4+5.6}{10}$

$\bar{x} = 4.1$

$\bar{x} = 4.1 \times 1000 = 4100$

We know that, $s^2 = \sum \dfrac{(x_i-\bar{x})^2}{n-1}$

x_i	$x_i - \bar{x}$	$(x_i-\bar{x})^2$
1.2	-2.9	8.41
4.6	0.5	0.25
3.9	-0.2	0.04
4.1	0	0
5.2	1.1	1.21
3.8	-0.3	0.09
3.9	-0.2	0.04
4.3	0.2	0.04
4.4	0.3	0.09
5.6	1.5	2.25
		12.42

$s^2 = \dfrac{\sum(x_i-\bar{x})}{n-1} = \dfrac{12.42}{9} = 1.38 \times 1000$

$= 1380$

$s = 37.148$

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{4100-4000}{37.148/\sqrt{10}}$$

$$\boxed{t = 8.5126}$$

Conclusion: $t_{cv} = 8.5126$

$t_{tv} = t_{\nu,\alpha/2} = t_{9,0.025} = 2.262$

$t_{cv} > t_{tv}$, so we are rejecting H_0 & accept H_1

$\therefore$ The average life time of bulb is not 4000 hrs.

Test of significance of different of mean's (small samples)

The test static for difference of mean's is given by

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

where $s_1 = \frac{\sum(x_i-\bar{x})^2}{n_1-1}$ $\quad s_2 = \frac{\sum(y_i-\bar{y})^2}{n_2-1}$

Degree's of freedom $(\nu) = n_1+n_2-2$

Below are given the gain in weights (Lvs) of pigs fed into diets A & B

Diet A 25 32 30 34 24 14 32 24 30 31 25 35 - - -

Diet B 44 34 22 10 47 31 40 30 32 35 18 21 35 29 22

Test if the 2 diets differ significantly as regarded their effect on the increase on weights.

Given $n_1 = 12$

$n_2 = 15$

Sample mean $\bar{x} = \frac{25+32+30+34+24+14+32+24+30+31+25+35}{12}$

$\bar{x} = 28$

$\bar{y} = 30$

Since 2 samples n_1, n_2 and their mean's $\bar{x}$ & $\bar{y}$ are available so we use test of significance for difference of means

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $s_1^2 = \frac{\Sigma (x_i - \bar{x})^2}{n_1 - 1}$ $\quad s_2^2 = \frac{\Sigma (y_i - \bar{y})^2}{n_2 - 1}$

Degree of freedom, $\nu = n_1 + n_2 - 2$

$= 12 + 15 - 2$

$\nu = 25$

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
25	-3	9
32	4	16
30	2	4
34	6	36
24	-4	16
14	-14	196
32	4	16
24	-4	16
30	2	4
31	3	9
25	-3	9
35	7	49
		380

$S_1 = \frac{\Sigma (x_i - \bar{x})^2}{n_1 - 1}$

$S_1 = \frac{380}{12-1}$

$= \frac{380}{11}$

$S_1 = 34.54$

y_i	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
44	14	196
34	4	16
22	-8	64
10	-20	400
47	17	289
31	1	1
40	10	100
30	0	0
32	2	4
35	5	25
18	-12	144
21	-9	81
35	5	25
29	-1	1
22	-8	64

$$\Sigma (y_i - \bar{y})^2 = 1410$$

$$S_2 = \frac{\Sigma (y_i - \bar{y})^2}{n_2 - 1}$$

$$= \frac{1410}{15-1}$$

$$= \frac{1410}{14}$$

$$S_2 = 100.71$$

Null hypothesis (H_0):

There is no significance btwn 2 diets

i.e., $\bar{x} = \bar{y}$

Alternative hypothesis (H_1):

There is significance btwn 2 diets

i.e, $\bar{x} = \bar{y}$

Level of significance:

$\alpha = 0.05$

Test for statistic:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} - \frac{s_2^2}{n_2}}} = \frac{28 - 30}{\sqrt{\frac{34.54^2}{12} - \frac{100.71^2}{15}}} = \frac{-2}{\sqrt{9.352}} = -0.6457$$

$$t_{cv} = 0.6457$$

$$t_{tv} = t_{\nu, \alpha/2}$$

$$\nu = n_1 + n_2 - 2 = 12 + 15 - 2 = 25$$

$$= t_{25, 0.025}$$

$$t_{tv} = 2.060$$

Conclusion:

since $t_{cv} < t_{tv}$ we accept null hypothesis

There is no significance differeence btwn the diets A & B.

(2) The blood pressure of 5 women before & after intake of a ceetain drug are given below

Before	110	120	125	132	125
After	120	118	125	136	121

Test whether there is significant change in blood pressure at 1% level.

~~Sol~~ Since the availability of differences btwn before and after we use test of significance, for paired data with differences, $d_1, d_2, \ldots d_n$ using test static as

$$t = \frac{\bar{d} - \mu}{S/\sqrt{n}} \Rightarrow t = \frac{\bar{d}}{S/\sqrt{n}} \quad (\because \mu = 0)$$

$$\bar{d} = \frac{\sum d_i}{n} \qquad S^2 = \sum \frac{(d_i - \bar{d})^2}{n-1}$$

Degree of freedom $(\nu) = n - 1$

We have $n = 5$ (sample - small)

Since the availability of differences (before and after)

we use $t = \frac{\bar{d}}{s/\sqrt{n}}$

where $\bar{d} = \frac{\sum d_i}{n}$

The difference d_i are $= \frac{-10+2+0-4+4}{5}$

$\bar{d}_i = -1.6$

where $s^2 = \sum\left(\frac{d_i - \bar{d}}{n-1}\right)^2$

d_i	$d_i - \bar{d}$	$(d_i - \bar{d})^2$
-10	-8.4	70.56
2	3.6	12.96
0	1.6	2.56
-4	-2.4	5.76
4	5.6	31.36
		123.2

$s^2 = \frac{123.2}{5-1}$

$s^2 = \sqrt{30.8}$

$= 5.55$

Null hypothesis: There is no significance change in the blood pressure

Alternative hypothesis: There is a significant change in the blood pressure.

It is TTT

Level of significance: $\alpha = 0.01$

Test of statistic: $t = \frac{-1.6}{5.55/\sqrt{5}}$

$= -0.645$

$|t| = 0.645$

Conclusion : $t_{cv} = 0.645$

$t_{TV} = t_{\nu, \frac{0.01}{2}} = t_{4, 0.005}$ $\quad \nu = n-1 = 5-1 = 4$

$t_{TV} = 4.604$

~~Com~~ $t_{CV} < t_{TV}$, we accept H_0

There is no significance change in the blood pressure.

③ To examine the hypothesis that the husbands are more intelligent than wife's, an investigator took a sample off couples and administrated then a test which measures IQ, the following results are

Husbands	117	105	97	105	123	109	86	78	103	107
Wives	106	98	87	104	116	95	90	69	108	85

Test the hypothesis with reasonable test at 5%.

④ Two independent samples 8 & 7 items respectively had the following values

Sample 1	11	11	13	11	15	9	12	14
Sample 2	9	11	10	13	9	8	10	-

Is the difference btwn mean of sample is significant.

Q: Memory capacity of 10 students were tested before and after training, state whether the training was effective or not from the following scores.

Before training	12	14	11	8	7	10	3	0	5	6
After training	15	16	10	7	5	12	10	2	3	8

Null hypothesis: The training was not effective before and after training.

Alternative hypothesis: The training was effective, before and after training

which is TTT

Level of significance: $\alpha = 0.05$

Test statistic:

Since the availability of before and after training so we use test of significance of paired samples given by

$$t = \frac{\bar{d}}{S/\sqrt{n}}$$

$n = 10$

The differences are,

$$\bar{d} = \frac{-3, -2, 1, 1, 2, -2, -7, -2, 2, -2}{10}$$

~~doee~~

$$\bar{d} = -1.2$$

where $s^2 = \sum \frac{(d_i - \bar{d})^2}{n-1}$

d_i	$d_i - \bar{d}$	$(d_i - \bar{d})^2$
-3	$-3+1.2 = -1.8$	3.24
-2	$-2+1.2 = -0.8$	0.64
1	$1+1.2 = 2.2$	4.84
1	2.2	4.84
2	3.2	10.24
-2	-0.8	0.64
-7	-5.8	33.64
-2	-0.8	0.64
2	3.2	10.24
-2	-0.8	0.64

$$\sum (d_i - d)^2 = 69.6$$

$$s^2 = \frac{69.6}{10-1} = \frac{69.6}{9} = 7.73$$

$$s = 2.78$$

$$t = \frac{\bar{d}}{s/\sqrt{n}} = \frac{-1.2}{2.78/\sqrt{10}} = -1.365$$

$$t_{cv} = -1.365$$

$$t_{tv} = t_{0.025,9} = 2.262$$

Conclusion:

$t_{cv} < t_{tv}$, so we accept null hypothesis.

The training was not effected, before & after training

Snedecor's F- distribution:

~~Since~~ When the availability of variances, we use Snedecor's F test given by,

$$F = \frac{S_1^2}{S_2^2} \text{ or } \frac{S_2^2}{S_1^2} \text{ where,}$$

$$S_1^2 = \frac{\Sigma (x_i - \bar{x})^2}{n_1 - 1} \qquad S_2^2 = \Sigma \frac{(y_i - \bar{y})^2}{n_2 - 1}$$

with degree's of freedom

$$\nu_1 = n_1 - 1 \quad , \quad \nu_2 = n_2 - 1$$

1) Determine (i) $F_{0.05}$ for $\nu_1 = 15$ $\nu_2 = 7 \rightarrow 3.51$

(ii) $F_{0.95}$ for $\nu_1 = 10$ $\nu_2 = 20$

(iii) $F_{0.001}$ for $\nu_1 = 6$ $\nu_2 = 22 \rightarrow 3.76$

(iv) $F_{0.99}$ for $\nu_1 = 6$ $\nu_2 = 20$

(ii) $F_{0.95} = \frac{1}{F_{0.05(20,10)}} = \frac{1}{2.77} = 0.3610$

(iv) $F_{0.99} = \frac{1}{F_{0.01(20,6)}} = \frac{1}{7.40} = 0.1351$

2) In one sample of 8 observations from a normal population, the sum of squares of deviations of the sample values from sample mean is 84.4, In another sample of 10 observations it was 102.6 Test at 5% level whether the populations have the same variance.

$n_1 = 8$ $\qquad\qquad$ $n_2 = 10$

$\sum(x_i - \bar{x})^2 = 84.4$ $\qquad$ $\sum(y_i - \bar{y})^2 = 102.6$

$$S_1^2 = \cancel{84.4}\, \frac{\sum(x_i - \bar{x})^2}{n_1 - 1} \qquad S_2^2 = \frac{\sum(y_i - \bar{y})^2}{n_2 - 1}$$

$$= \frac{84.4}{7} \qquad\qquad = \frac{102.6}{9}$$

$$S_1^2 = 12.0571 \qquad\qquad S_2^2 = 11.4$$

$$\nu_1 = n_1 - 1 = 8 - 1 = 7 \qquad\qquad \nu_2 = n_2 - 1 = 10 - 1 = 9$$

Null hypothesis (H_0):

The populations have the same valiances

$$S_1^2 = S_2^2$$

Alternative hypothesis (H_1):

The populations doesnot hv the same vaeiances

$$S_1^2 \neq S_2^2$$

which is TTT (Two Tailed Test)

Level of significance:

$$\alpha = 5\% = 0.05$$

Test statistic:

$$F = \frac{S_1^2}{S_2^2}$$

$$= \frac{12.0571}{11.4}$$

$$F_{cv} = 1.0576$$

$$F_{TV} = F_{0.05}(\nu_1, \nu_2)$$

$$= F_{0.05}(7, 9)$$

$$F_{TV} = 3.29$$

$\therefore$ $F_{cv} < F_{TV}$, we accept H_0 (Null hypothesis)

The populations hv the same variances

$$S_1^2 = S_2^2$$

(3) The following random samples are the measurements of heat producing capacity (in millions of calaries per ton) of specimens of coal from 2 mines

Mine 1	8260	8130	8350	8070	8340	-
Mine 2	7950	7890	7900	8140	7920	7840

Use 0.02 level of significance to test whether it is reasonable to assume that the variances of 2 populations are equal.

(4) The nicotin contents in milligrams in 2 samples of tobacco were found to be as follows

Sample A	24	27	26	21	25	-
Sample B	27	30	28	31	22	36

Can it be said that the 2 samples have come from the same normal population

Sol: To test whether the sample have come from same normal population, we use

(i) Equality of variances by F-test $\left[F = \frac{S_1^2}{S_2^2} \text{ or } \frac{S_2^2}{S_1^2} \right]$

(ii) Equality of mean's by T-test $\left[T = \frac{\bar{x} - \bar{y}}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \right]$

Method - 1:

(i) To test equality of variance by F-Test

Null hypothesis

The variances have come from same normal population

(or)

The 2 variances are equal.

$$S_1^2 = S_2^2$$

Alternative hypothesis:

The variances doesnot come from same normal population

$$S_1^2 \neq S_2^2$$

which is TTT

Level of significance

$$\alpha = 0.05$$

$$\bar{x} = \frac{24+27+26+21+25}{5}$$

$$\bar{x} = 24.6$$

Test statistic

$$\bar{y} = \frac{27+30+28+31+22+36}{6}$$

$$F = \frac{S_1^2}{S_2^2} \text{ or } \frac{S_2^2}{S_1^2}$$

$$\bar{y} = 29$$

$$S_1^2 = \frac{\Sigma(x_i - \bar{x})^2}{n_1 - 1} \qquad S_2^2 = \frac{\Sigma(y_i - \bar{y})^2}{n_2 - 1}$$

$$\nu_1 = n_1 - 1 = 5 - 1 = 4 \qquad \nu_2 = n_2 - 1 = 6 - 1 = 5$$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
24	-0.6	0.36
27	2.4	5.76
26	1.4	1.96
21	-3.6	12.96
25	0.4	0.16

$\Sigma(x_i - \bar{x})^2 = 21.2$

y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
27	2	4
30	1	1
28	1	1
31	2	4
22	7	49
36	7	49

$\Sigma(y_i - \bar{y})^2 = 108$

$$S_1^2 = \frac{21.2}{4} = 5.3 \qquad S_2^2 = \frac{108}{5} = 21.6$$

$= \cancel{4.024}$

$= 4.206$

$$\therefore F = \frac{S_2^2}{S_1^2} = \frac{21.6}{5.3} = 4.0755$$

Conclusion: $F_{CV} = 4.0755$

$$F_{TV} = F_{0.05}(\nu_1, \nu_2) = F_{0.05}(4,5) = 5.19$$

$F_{CV} < F_{TV}$, so we accept H_0

The variances have came from same normal population

$$S_1^2 = S_2^2$$

Method 2:

(ii) To test equality of mean's by T-test

Null hypothesis: The mean's have come from same normal population

$$\bar{x} = \bar{y}$$

Alternative hypothesis: The mean's does not come from same normal population

$$\bar{x} \neq \bar{y}$$

which is TTT

Level of significance: $\alpha = 0.05$

Test statistic:

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{24.6 - 29}{\sqrt{\frac{5.3}{5} + \frac{21.6}{6}}}$$

$$= \frac{-4.4}{\sqrt{1.06 + 3.6}} = \frac{-4.4}{\sqrt{4.66}} = \frac{-4.4}{2.1587} = -2.0383$$

Conclusion: $T_{cv} = 2.0883$

T_{TV} $\nu = n_1 + n_2 - 2$

$\nu = 5 + 6 - 2$

$\nu = 9$

$T_{TV}(\nu, \alpha/2)$

$T_{TV}(9, 0.025) = 2.262$

$T_{cv} < T_{TV}$, so we accept H_0

The mean's hv come from same normal population

$\bar{x} = \bar{y}$

$\therefore$ By variance and mean's test's we conclude that, the 2 samples hv come from same normal population

(3) sol:

Null hypothesis

The 2 variances are equal

$$S_1^2 = S_2^2$$

Alternative hypothesis:

The 2 variances are not equal

$$S_1^2 \neq S_2^2$$

which is TTT

Level of significance

Consider standard level of significance, $\alpha = 0.05$

Test statistic

$$F = \frac{S_1^2}{S_2^2} \text{ or } \frac{S_2^2}{S_1^2}$$

$$S_1^2 = \frac{\sum(x_i-\bar{x})^2}{n_1-1} \qquad S_2^2 = \frac{\sum(y_i-\bar{y})^2}{n_2-1}$$

$\nu_1 = n_1 - 1$

$= 5-1$

$\nu_1 = 4$

$\nu_2 = n_2 - 1$

$= 6-1$

$\nu_2 = 5$

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
8260	30	900
8130	-100	10000
8350	120	14400
8070	-160	25600
8340	110	12100

$\sum(x_i-\bar{x})^2 = 63000$

y_i	$y_i-\bar{y}$	$(y_i-\bar{y})^2$
7950	10	100
7890	-50	2500
7900	-40	1600
8140	200	40000
7920	-20	400
7840	-100	10000

$\sum(y_i-\bar{y})^2 = 54600$

$$\bar{x} = \frac{8260+8130+8350+8070+8340}{5}$$

$\bar{x} = 8230$

$$\bar{y} = \frac{7950+7890+7900+8140+7920+7840}{6}$$

$\bar{y} = 7940$

$$S_1^2 = \frac{63000}{4}$$

$S_1^2 = 15750$

$$S_2^2 = \frac{54600}{5}$$

$S_2^2 = 10920$

$$F = \frac{S_1^2}{S_2^2} = \frac{15750}{10920} = 1.4423$$

Conclusion: $F_{cv} = 1.4423$

$F_{Tv} = F_{0.05}(\nu_1, \nu_2) = F_{0.05}(4,5) = 5.19$

$F_{cv} < F_{Tv}$, so we accept H_0

the 2 variances hence are equal

$S_1^2 = S_2^2$

Chi-Square Distribution (χ^2)

If O_i is the set of observed frequencies and E_i are the expected frequency or theoritical frequencies, the χ^2 is defined as

$$\chi^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

Conditions to apply χ^2 test:

1. The sample of observations should be independent.
2. The total frequency is large (>50)
3. No theoritical frequency should be less than 10 (<10).

Method-1: Test of significance for goodness of fit
(or)
χ^2 test of goodness of fit

→ Test statistic is given by

$$\chi^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

~~The number~~

Degrees of freedom: $\nu = n-1$

Q: The no. of automobile accidents per week in a certain community are as follows.

12, 8, 20, 2, 14, 10, 15, 6, 9, 4,

are these frequencies in agreement with the belief that the accident conditions were same during this 10 week period.

$n = 10$, it is a small sample.

The observed frequencies are O_i :

12 8 20 2 14 10 15 6 9 4

Expected frequency (E_i):

$$E_1 = \frac{10+8+20+2+14+10+15+6+9+4}{10} = \frac{100}{10} = 10$$

Since observed frequencie's and E_i is availabe so we use test of significance for goodness of fit. given by

$$\chi^2 = \Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

Degrees of freedom (D.O.F) = $\nu = n-1 = 10-1$

$\nu = 9$

Null hypothesis (H_0): The accidents conditions were ~~not~~ same during this 10 week period.

Alternative Hypothesis (H_1): The accidents conditions were not same during this 10 week period

Level of significance: $\alpha = 0.05$

Test statistic is $\chi^2 = \Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right]$

O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
12	10	2	4	0.4
8	10	-2	4	0.4
20	10	10	100	10
2	10	-8	64	6.4

14	10	4	16	1.6
10	10	0	0	0
15	10	5	25	2.5
6	10	4	16	1.6
9	10	1	1	0.1
4	10	6	36	3.6

$$\sum\left(\frac{(O_i-E_i)^2}{E_i}\right) = 26.6$$

Conclusion:

$$\chi^2_{CV} = 26.6$$

$$\chi^2_{TV} = \chi^2_{\nu,\alpha} = \chi^2_{9,0.05} = 16.919$$

$\chi^2_{CV} > \chi^2_{TV}$, So we will not accept H_0 ~~&~~ accept H_1

The accidents conditions were not same during this 10 week period.

② The sample analysis of examination results of 500 students was made, it was found that 220 were failed, 170 got 3rd class, 90 placed in 2nd class and 20 got a 1st class. Do these figures communsurate with the general examination results which are in the ratio 4:3:2:1 for the various categories respectively

$n = 4$

The observed frequencies (O_i):

220 170 90 20

Null hypothesis (H_0): These figures communsurate with the general examination results.

Alternative Hypothesis (H_1): These figures ~~does~~ are not communsurate with general examination results which is TTT

Level of significance: $\alpha = 0.05$ (standard)

Test statistic:

$$\chi^2 = \Sigma \left(\frac{(O_i - E_i)^2}{E_i} \right)$$

O_i : 220, 170, 90, 20

Expected frequencies E_i are:

$E_i(\text{failed}) = E_i(220) = 500 \times \frac{4}{10} = 200$

$E_i(\text{Ist class}) = E_i(170) = 500 \times \frac{3}{10} = 150$

$E_i(\text{II}) = E_i(90) = 500 \times \frac{2}{10} = 100$

$E_i(\text{III}) = E_i(20) = 500 \times \frac{1}{10} = 50$

category	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
Fail	220	200	20	400	2
III	170	150	20	400	2.66
II	90	100	-10	100	1
I	20	50	-30	900	18

$$\Sigma \left(\frac{(O_i - E_i)^2}{E_i} \right) = 23.66$$

$\therefore \chi^2 = 23.66$

Conclusion:

$$\chi^2_{cv} = 23.66$$

$$\nu = n-1 = 4-1 = 3$$

$$\chi^2_{TV} = \chi^2_{\nu,\alpha} = \chi^2_{3,0.05}$$

$$= 7.815$$

since $\chi^2_{cv} > \chi^2_{TV}$ so we reject H_0 & accept H_1

These figures does not communsuerate with general examination results.

(3) Given are the frequencies of the digits from a telephone directery

Digits	0	1	2	3	4	5	6	7	8	9
frequencies	12	8	4	6	11	10	9	13	12	18

Use χ^2-test to indentify, that the frequencies are significant at 1% level.

Given $n = 10$

Null hypothesis (H_0): The frequencies are significant

Alternative hypothesis (H_1): The frequencies are not significant which is TTT

Level of significance: $\alpha = 0.01$

Test statistic:

$$\chi^2 = \Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

Expected frequencies (E_i):

$$E_i = \frac{12+8+4+6+11+10+9+13+12+18}{10} = \frac{103}{10}$$

$$E_i = 10.3$$

Observed frequencies (O_i):

12, 8, 4, 6, 14, 10, 9, 13, 12, 18

O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
12	10.3	1.7	2.89	0.2806
8	10.3	-2.3	5.29	0.5136
4	10.3	-6.3	39.69	3.8534
6	10.3	-4.3	18.49	1.7951
11	10.3	0.7	0.49	0.0476
10	10.3	-0.3	0.09	0.0087
9	10.3	-1.3	1.69	0.1640
13	10.3	2.7	7.29	0.7078
12	10.3	1.7	2.89	0.2806
18	10.3	7.7	59.29	5.7563

$$\Sigma \left[\frac{(O_i - E_i)^2}{E_i} \right] = 13.4078$$

$$\chi^2 = 13.40$$

Conclusion:

$$\chi^2_{CV} = 13.40$$

$$\nu = n-1 = 10-1 = 9$$

$$\chi^2_{TV} = \chi^2_{\nu,\alpha} = \chi^2_{9,\,0.01} = 21.666$$

since, $\chi^2_{CV} < \chi^2_{TV}$, we accept H_0

The frequencies are significant.

(9) A pair of fair dice ~~are~~ thrown 360 times & frequency of each sum is indicated below

Sum	2	3	4	5	6	7	8	9	10	11	12
frequency	8	24	35	37	44	65	51	42	26	14	14

Would you say that the dice are fair on the basis of χ^2 test at 5% level of significance.

$n = 1\phi$

Null Hypothesis: The dice are fair

Alternative Hypothesis: The dice are not fair

Level of significance: $\alpha = 0.05$

Test statistic:

$$\chi^2 = \sum\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

(observed frequencies)

O_i's are 8, 24, 35, 37, 44, 65, 51, 42, 26, 14, 14

To evaluate E_i's;

~~E_R(Expected frequencies)~~

The probabilities of getting sum, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 are

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

Sums	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
2	8	$360 \times \frac{1}{36} = 10$	-2	4	0.4
3	24	$360 \times \frac{2}{36} = 20$	4	16	0.8
4	35	$360 \times \frac{3}{36} = 30$	5	25	0.8 33

5	37	$360 \times \frac{4}{36} = 40$	-3	9	0.225
6	44	$360 \times \frac{5}{36} = 50$	-6	36	0.72
7	65	$360 \times \frac{6}{36} = 60$	5	25	0.4166
8	51	$360 \times \frac{5}{36} = 50$	1	1	0.02
9	42	$360 \times \frac{4}{36} = 40$	2	4	0.1
10	26	$360 \times \frac{3}{36} = 30$	-4	16	0.533
11	14	$360 \times \frac{2}{36} = 20$	-6	36	1.8
12	14	$360 \times \frac{1}{36} = 10$	-4	16	1.6

$$\Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right] = 7.448$$

Conclusion: $\chi^2_{CV} = 7.448$

$\chi^2_{TV} = \chi^2_{\nu,\alpha}$ $\quad \nu = 11 - 1 = 10$

$= \chi^2_{10, 0.05}$

$\chi^2_{TV} = 18.307$

$\chi^2_{CV} < \chi^2_{TV}$, H_0 is accepted

$\therefore$ The dice are fair

③ Fit a poison distribution to the following data and its goodness of fit at level of significance = 0.05.

x	0	1	2	3	4
frequency (f)	419	352	154	56	19

Null Hypothesis(H_0):

Poison distribution is fitted to the given data

Alternative Hypothesis(H_1):

poison distribution is not fitted to the given data

Level of significance: $\alpha = 0.05$

Test statistic:

$$\chi^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

The O_i's are: 419, 352, 154, 56, 19

E_i's are to be evaluated using poison distribution

given by. $N \cdot \frac{e^{-\lambda} \lambda^x}{x!}$

where $N = \sum f_i = 419 + 352 + 154 + 56 + 19 = 1000$

$$\text{Mean}, \mu = \frac{\sum f_i x_i}{\sum f_i} = \frac{0 \times 419 + 1 \times 352 + 2 \times 154 + 3 \times 56 + 4 \times 19}{1000}$$

$$= \frac{904}{1000} = 0.904$$

Mean of poison distribution, $\mu = \lambda = 0.904$

Poison frequencies are,

$$N \cdot P(x) = N \left[\frac{e^{-\lambda} \lambda^x}{x!} \right]$$

when $x = 0 \Rightarrow 1000 \left[\frac{e^{-0.904} (0.904)^0}{0!} \right] = 404.9466$

$x = 1 \Rightarrow 1000 \left[\frac{e^{-0.904} (0.904)^1}{1!} \right] = 366.0718$

$$x = 2 \Rightarrow 1000\left[\frac{e^{-0.904}(0.904)^2}{2!}\right] = 165.4644$$

$$x = 3 \Rightarrow 1000\left[\frac{e^{-0.904}(0.904)^3}{3!}\right] = 49.8599$$

$$x = 4 \Rightarrow 1000\left[\frac{e^{-0.904}(0.904)^4}{4!}\right] = (405.6605) \times$$

11.2683

x	0	1	2	3	4
$f(x)$	419	352	154	56	19
p. $f(x)$	405	366	165	50	11

x	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
0	419	405	14	196	0.4840
1	352	366	-14	196	0.5355
2	154	165	-11	121	0.7333
3	56	50	6	36	0.72
4	19	11	8	64	5.8182
				$\sum \frac{(O_i - E_i)^2}{E_i}$	8.291

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 8.291$$

Conclusion

$\chi^2_{cv} = 8.291$

$\chi^2_{Tv} = \chi^2_{\nu, \alpha}$

$\chi^2 = \chi^2_{4, 0.05}$

$= 9.488$

$\chi^2_{cv} < \chi^2_{Tv}$

we accept H_0

Poison distribution is fitted to given data

Test of significance for independent attributes

The test statistic for independent attribute is given by

$$\chi^2 = \Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

with degree's of freedom, $\nu = (\text{no. of rows} - 1)(\text{no. of columns} - 1) = (r-1)(c-1)$

Expected frequencies (E_i's) to be calculated for contigensy table as

	A	B	
X	$E(a) = \frac{(a+b)(a+c)}{N}$	$E(b) = \frac{(a+b)(b+d)}{N}$	$a+b$
y	$E(c) = \frac{(c+d)(a+c)}{N}$	$E(d) = \frac{(c+d)(b+d)}{N}$	$c+d$
	$a+c$	$b+d$	$a+b+c+d = N$

1Q: 1000 students at college level were graded according to their IQ's and the economic conditions of their home, use χ^2-test to find whether their is any association btwn the economic conditions at home & IQ's of the students

Economic condition	IQ high	IQ low	total
Rich	460	140	600
Poor	240	160	400
Total	700	300	1000

Null Hypothesis (H_0): There is no association btwn the economic conditions at home & IQ's of the students

Alternative hypothesis (H_1): There is a association btwn the economic conditions at home & IQ's of the students

Level of significance which is TTT $\alpha = 0.05$ (standard)

Test statistic:

$$\chi^2 = \Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

O_i's are 460, 140, 240, 160

Expected frequencies are,

$$E(460) = \frac{600 \times 700}{1000} = 420$$

$$E(140) = \frac{600 \times 300}{1000} = 180$$

$$E(240) = \frac{700 \times 400}{1000} = 280$$

$$E(160) = \frac{400 \times 300}{1000} = 120$$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
460	420	40	1600	3.8095
140	180	-40	1600	8.888
240	280	40	1600	5.7142
160	120	-40	1600	13.3333

$$\chi^2 = \Sigma\frac{(O_i - E_i)^2}{E_i} = 31.7458$$

$$\chi^2 = 31.7458$$

Conclusion: $\chi^2_{cv} = 31.7458$

$$\chi^2_{TV} = \chi^2_{\upsilon,\alpha}$$

$$\upsilon = (r-1)(c-1) = (2-1)(2-1) = 1$$

$$\chi^2_{TV} = \chi^2_{1,0.05}$$

$$\chi^2_{TV} = 3.841$$

$\chi^2_{cv} > \chi^2_{TV}$ so we reject Null hypothesis (H_0) & accept H_1

There is association btwn economic conditions at home & IQ's of student

② Q: 4 methods are under development for making disc's of a super conducting material. 50 disc's are made by each method & they are checked for the super conductivity when cooled with liquid.

	Methods 1st	2nd	3rd	4th	
Super conductors	31	42	22	25	120
Failures	19	8	28	25	80
	50	50	50	50	200

Test the significance difference btwn the proportions of super conductors at 0.05 level

(H_0) Null Hypothesis: There is no difference btwn the proportion of super conductors

(H_1) Alternative Hypothesis: There is a difference btwn the proportion of super conductors

which is TTT

Level of Significance: $\alpha = 0.05$ (given)

Test statistic:

$$\chi^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i}\right]$$

O_i's are 31, 42, 22, 25, 19, 18, 28, 25

Expected frequencies (E_i's) are

$E(31) = \frac{50 \times 120}{200} = 30$ $E(19) = \frac{80 \times 50}{200} = 20$

$E(42) = \frac{50 \times 120}{200} = 30$ $E(8) = \frac{80 \times 50}{200} = 20$

$E(22) = \frac{50 \times 120}{200} = 30$ $E(28) = \frac{80 \times 50}{200} = 20$

$E(25) = \frac{50 \times 120}{200} = 30$ $E(25) = \frac{80 \times 50}{200} = 20$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
31	30	1	1	0.033
42	30	12	144	4.8
22	30	-8	64	2.133
25	30	-5	25	0.8333
19	20	-1	1	0.05
8	20	-12	144	7.2
28	20	8	64	3.2
25	20	5	25	1.25

$$\sum \frac{(O_i - E_i)^2}{E_i} = 19.4996$$

$$\chi^2_{cv} = 19.4996$$

Conclusion: $\chi^2_{CV} = 19.4996$

$$\chi^2_{TV} = \chi^2_{\nu, \alpha}$$

$$\nu = (r-1)(c-1) = (2-1)(4-1) = (1)(3)$$

$$\chi^2_{TV} = \chi^2_{3, 0.05} = 7.815$$

$\chi^2_{CV} > \chi^2_{TV}$ so we reject H_0 (Null Hypothesis) and we accept H_1

There is a difference btwn the proportions of super conductors

3Q: 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

	Favourable	Not Favourable	
New	60	30	90
Conventional	40	70	110
	100	100	200

Null Hypothesis (H_0): The new & conventional treatments are same

Alternative Hypothesis (H_1): The new & ~~conventional~~ treatment is ~~not~~ ~~same~~ superior to conventional treatment which is RTT

Level of significance: $\alpha = 0.05$

Test Statistic:

$$\chi^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i}\right]$$

The O_i's are 60, 30, 40, 70

$$E(60) = \frac{100 \times 90}{200} = 45$$

$$E(30) = \frac{100 \times 90}{200} = 45$$

$$E(40) = \frac{110 \times 100}{200} = 55$$

$$E(70) = \frac{110 \times 100}{200} = 55$$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
60	45	15	225	5
30	45	15	225	5
40	55	15	225	4.0909
70	55	15	225	4.0909

$$\sum \frac{(O_i - E_i)^2}{E_i} = 18.1818$$

$$\chi^2_{CV} = 18.1818$$

Conclusion: $\chi^2_{CV} = 18.1818$

$\chi^2_{TV} = \chi^2_{\nu, \alpha}$

$= \chi^2_{1, 0.05}$

$\chi^2_{TV} = 3.841$

$\nu = (r-1)(c-1)$

$= (2-1)(2-1)$

$= 1$

$\chi^2_{CV} > \chi^2_{TV}$, so we reject Null hypothesis (H_0) & accept Alternative Hypothesis.

The new treatment is superior to conventional treatment.

STOCHASTIC PROCESS

In some situations in science & technology we will be interested in studying "**Processes**", i.e., the phenomena that takes place with change in time. The theory of probability studied did not have either general procedure or elaborate schemes for solving problems that arise in the study of such phenomena. Hence it is necessary to develop a general theory of **random process** or **stochastic process** to study random variables dependent on one or several discrete or continuous varying parameters.

Mathematically, a **stochastic process** is a set of random variables **{X_t}** or **X{t}** depending on some real parameter like time t, these are also known as **random process or random functions**.

The family of random variables which are the functions of time parameter is known as **stochastic process or random process**.

The **stochastic process** is a family of random variables **{ X(t) / t ∈ T** } is defined on a given probability space indexed by time parameter **'t'**.

The values assumed by the process are called '**states**' and the set of all possible values of an individual member of the random process is called the '**state space**', which can be either discrete or continuous.

Examples of stochastic process:

Queuing system

In queuing system, the number of people joining the queue during the time interval, the number of people served from the queue in particular time interval.

Turbulent fluid flow

In the turbulent fluid flow, the velocity components **u, v, w** of the fluid are random variables depending on the space coordinates **x, y, z** and the time.

Movement of molecules of a gas or liquid:

In a gas or liquid at random instants the molecules collides with the other molecules. Thus its velocity and position are altered. Thus the state of the molecule is subjected to random changes at every instant of time.

Communication process

The amplitude of the signals to be transmitted and amplitude of the noise produced in the channel depending on the time are both random variables.

A random walk model

A particle moves in a straight line in steps of unit length. At each stage it can move one step to the right with probability **p** or one step to the left with probability **q.**

If the particle starts from origin, its position after **n** movements is a random variable, which depends on discrete parameter **n.**

Gamblers Ruin Problem (Classical Ruin Problem)

Suppose a Gambler wins or loses a unit (Rupee/Dollar/Pound) with probabilities **p & q** respectively.

Let his initial capital be **z** & his opponents initial capital be **(a-z)**, so that the combined capital is **a.**

The game continuous until the Gambler's capital is either reduced to **zero** or increased to **a.** i.e., until one of the two players is ruined. Here we notice that the capital of the Gambler after **n** stages is a random stochastic variable.

Classification of stochastic process/random process:

A random process can classified according to the characteristics of **t** and the random variable **X(t)**.

(i) Discrete random process:

A random process **X(t,s)** which corresponds to the random variable **X** having only discrete values while **t** is continuous, is known as discrete random process.

If **x** is continuous and **t** is discrete, is known as discrete random process.

It is represented by $\{X_n ; n=1,2,..\}$

(ii) Continuous random process:

If **X** is continuous and **t** is any of continuous values, then **X(t)** is called continuous random process. The process is defined for all the instants of time as $t \geq 0$, then it is said to be continuous stochastic process and is represented by $\{X(t) ; t \geq 0\}$

Deterministic and non-deterministic stochastic process:

If the future values of any sample function cannot be predicted exactly from the observed values, the process is called a **non-deterministic process**.

A process is called **deterministic** if the future values of any sample function can be predicted from the past values.

A deterministic random process is also known as a **predictable random process**.

If the future values of any sample function cannot be predicted exactly from the past values, then it is called a **non-deterministic random process.** It is also called a **regular random process.**

Stationary and independent process:

For a **stationary stochastic process**, if all its statistical processes or the distribution function or certain expected values do not change with time.

Contradictorily, for a **non-stationary stochastic process**, any of its density functions or probability functions or any of its movements depends on the precise value of time.

A process which is not stationary is said to be **evolutionary**.

Normal or Gaussian stochastic process:

If the joint distribution of any number of variables $Xt_1, Xt_2, \ldots\ldots$ in normal distribution then it is said to be normal or Gaussian stochastic process.

Markov Process:

A stochastic process is said to be a Markov process or Markovian if,

$$P[X(t_{n+1}) \leq X_{n+1} / X(t_n)=X_n , X(t_{n-1}, \ldots, X(t_0)= t_0]=P[X(t_{n+1})= X(t_n)]=X_n \text{ where } t_0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq t_{n+1}$$

Here $X_0\ X_1\ X_2 \ldots X_z$ are called the states of the process.

Markov chain:

Let **X(t)** be a Markov process. If **X(t)** possesses Markov property and takes only discrete values, whether **t** is discrete or continuous, then **X(t)** is called Markov chain.

A sequence of states $[X_n]$ is Markov chain if each X_n is a random variable and if

$$P[X(t_{n+1})=X_{n+1} / X_n=x_n , X_{n-1}=x_{n-1} , \ldots X_0=x_0]=P[X_n=n/X_n=x_n]$$

Examples of Markov chains are,

A random walk on the line,

Random placements of balls.

Probability vector:

A vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ is called a probability vector if the components are **non-negative** and their **sum is equal** to **1**.

Transition probability matrix (tpm): The transition probabilities, P_{ij}'s are arranged in a matrix given by, $P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ \vdots & & \dots & \vdots \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{bmatrix}$

$$\left\{ \begin{array}{c} {*}{*}{*}\ \textit{Stochastic matrix when} \dots \\ i)\ \textit{square matrix} \\ ii)\ \textit{non-negative} \\ iii)\ \textit{each row sum is } 1 \quad {*}{*}{*} \end{array} \right\}$$

Stochastic matrix:

A square matrix $P=P_{ij}$ is called a stochastic matrix if **each of the row is a probability vector**, if each entry **P** is **non-negative** and the **sum of entries in each row** is **1**.

A stochastic matrix is a random matrix with **non-negative** elements and **unit row sums**.

A tpm is said to be a stochastic matrix if the elements of **each rows are non-negative** and **the sum of the elements in each rows is equal** to**1**.

A stochastic matrix **P** is said to be regular if all the entries of some power P^n are **positive**.

A stochastic matrix **P** is said to be not regular if **1** occurs in the principle diagonal.

A transition probability matrix (tpm), P is said to be regular matrix if all entries P^n **(n=1,2,...)** are **non-zero positive** values.

If every state can be reached from any state (in any number of transitions) then the chain is said to be **irreducible**. Otherwise, the chain is said to be **reducible** or **non-irreducible**.

A state **i** is said to be an **absorbing state** iff $P_{ij}=1$. A Markov chain is **absorbing** if it has at least one absorbing state and it is possible to go from every non-absorbing state to at least one absorbing state in one or more steps.

Ergodicity is restrictive forms of stationary processes from which the time and ensemble averages are interchangeable are called **ergodic process.**

A state is said to be **periodic** with period **t (>1)** if the return to the state is possible only in **t, 2t, 3t, 4t, ...** steps, where t is greatest integer with this property. In this case $P_{ij}^{(n)}=0$, unless **n** is an integral multiple of **t**. the state **i** is said to be **aperiodic (or non-periodic)** of no such **t(>1)** exists.

A positive recurrent (or positive persistent) and apriodic state is called **ergodic**. A Markov chain all of whose states are **ergodic** is said to be **ergodic chain**.

PROBLEMS ON STOCHASTIC PROCESS

- ✓ Which of the following matrices are **stochastic matrices**?

1. $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$
   Since the given matrix is not a square matrix, which is **not a stochastic matrix.**

2. $B = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$

The matrix is square matrix
The all entries in the matrix are positive
Each row sum in the matrix is equal to 1
Hence the given matrix is **stochastic matrix.**

3. $D = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 1/6 & 1/3 \\ 1/3 & 2/3 & 0 \end{bmatrix}$

The matrix is square matrix
The entries in the matrix are positive
Each row sum in the matrix is equal to 1
Hence the given matrix is **stochastic matrix.**

4. $E = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

The matrix is square matrix
The entre in the second row is a negative
Sum of each row is also not equal to 1
Hence the matrix is **not a stochastic matrix.**

5. $F = \begin{bmatrix} 0 & 2 \\ 1/4 & 1/4 \end{bmatrix}$

The matrix is square matrix
The entries in the matrix are positive
Each row sum in the matrix is not equal to 1
Hence the given matrix is **not a stochastic matrix.**

6. $G = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$

The matrix is square matrix
The entries in the matrix are positive
Each row sum in the matrix is equal to 1
Hence the given matrix is **stochastic matrix.**

7. $H = \begin{bmatrix} 1 & 0 \\ 1/3 & 2/3 \end{bmatrix}$, the matrix is **a stochastic matrix.**

8. $I = \begin{bmatrix} 15/16 & 1/16 \\ 2/3 & 4/3 \end{bmatrix}$, the matrix is **not a stochastic matrix.**

9. $J = \begin{bmatrix} 1/3 & 2/3 & 4/3 \\ 1/2 & 1 & 1/2 \end{bmatrix}$, the matrix is **not a stochastic matrix.**

Which of the following matrices are **regular matrices**?

1. $A = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$

The matrix is square matrix
The entries in the matrix are positive
Each row sum in the matrix is equal to 1
Hence the given matrix is **stochastic matrix.**

$$A^2 = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 1/4 \end{bmatrix}$$

$$A^3 = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 0.25 & 0.75 \\ 0.375 & 0.625 \end{bmatrix}$$

$$A^4 = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 0.375 & 0.625 \\ 0.3125 & 0.6875 \end{bmatrix}$$

.
.
.
.

Here each entry in **$A^2, A^3, A^4, \ldots,$** are **positive**, **Sum of all entries are equal to 1** & **no unit element in principle diagonal**.

Hence, **A** is a **regular matrix**.

2. $B = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}$

The matrix is square matrix
The entries in the matrix are positive
Each row sum in the matrix is equal to 1
Hence the given matrix is **stochastic matrix.**
The given matrix is **not regular,** since 1 lies on the principle diagonal.

3. $C = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$

The matrix is square matrix
The entries in the matrix are positive
Each row sum in the matrix is equal to 1
Hence the given matrix is **stochastic matrix.**

$$C^2 = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}\begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 1/2 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 3/8 & 3/8 & 1/4 \end{bmatrix}$$

$$C^3 = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}\begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}\begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 1/2 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 7/16 & 7/16 & 1/8 \end{bmatrix}..$$

Since, some of the entries in $C^2, C^3, \ldots$ are **zeros** & **not reduced** *****(zeros must be reduces, compare to next powers of matrices) *****

Hence, **C** is a **not regular matrix**.

4. $D = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$

The matrix is square matrix
The entries in the matrix are positive
Each row sum in the matrix is equal to 1

Hence the given matrix is **stochastic matrix.**

$$D^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix}$$

$$D^3 = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 1/4 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}..$$

$$D^4 = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/2 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}$$

$$D^5 = \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/8 & 1/4 & 3/8 \\ 1/4 & 1/4 & 1/2 \end{bmatrix} \ldots\ldots\ldots$$

Since, all the entries of some power of matrix D are positive, no unit element in principle diagonal & number of zeros also reducing compare to the powers of matrices.

Hence, **D** is a **regular matrix**.

5. $E = \begin{bmatrix} 1/2 & 1/4 & 1 \\ 0 & 1/2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$, the matrix is **not a stochastic matrix**; it is **not a regular matrix.**

6. $F = \begin{bmatrix} 1/3 & 0 \\ 1/3 & 1 \end{bmatrix}$, the matrix is **not a stochastic matrix**; it is **not a regular matrix.**

7. $G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, the matrix is **a stochastic matrix**. Since, **unit element in main diagonal**, so it is **not a regular matrix.**

The transition probability matrix (tpm)is given by $\begin{bmatrix} \mathbf{0.3} & \mathbf{0.7} & \mathbf{0} \\ \mathbf{0.1} & \mathbf{0.4} & \mathbf{0.5} \\ \mathbf{0} & \mathbf{0.2} & \mathbf{0.8} \end{bmatrix}$. Is the matrix is irreducible?

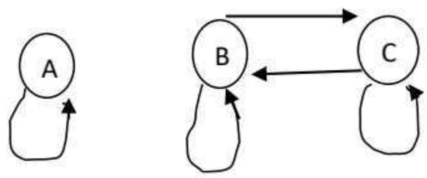
Consider the states 0, 1, 2
Here, it is possible to go from 0 to 1 with probability 0.7
And from state 1 to 2 with probability 0.5
Thus it is possible to go from 0 to 2.
In this way, all the states are communicate with each other
Hence, the chain is irreducible.

The transition probability matrix (tpm) is given by $\begin{bmatrix} \mathbf{0.4} & \mathbf{0.6} & \mathbf{0} & \mathbf{0} \\ \mathbf{0.3} & \mathbf{0.7} & \mathbf{0} & \mathbf{0} \\ \mathbf{0.2} & \mathbf{0.4} & \mathbf{0.1} & \mathbf{0.3} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}$. Is the matrix is irreducible?

Consider the states 0, 1, 2, 3
Here, we observe that from state 3 no other state is accessible
And principle diagonal contains 1 in state 3.
All the states are do not communicate with each other.
Hence, the chain is not irreducible.

Is the transition probability matrix (tpm), P = $\begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0.5} & \mathbf{0.5} \\ \mathbf{0} & \mathbf{0.33} & \mathbf{0.67} \end{bmatrix}$ an absorbing markov chain?

We have P = $\begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0.5} & \mathbf{0.5} \\ \mathbf{0} & \mathbf{0.33} & \mathbf{0.67} \end{bmatrix}$



Here, state A is only absorbing state and A cannot be reached from state B or state C.
Hence, P is not an absorbing markov chain.

A training process is considered as a two state markov chain. If it rains, it is considered to be in state 0 and it does not rains, the chain is in the state of 1.
The transition probability matrix (tpm), P = $\begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$.
Find the probability that it will rain for 3 days from today assuming that it is raining today.

The 1 step transition probability matrix (tpm) is given by, P = $\begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$

P(1 day) = P^1 = $\begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$

P(2 days) = P^2 = $\begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}\begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.44 & 0.56 \\ 0.28 & 0.72 \end{bmatrix}$

P(3 days) = P^3 = $\begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}\begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}\begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.376 & 0.624 \\ 0.312 & 0.688 \end{bmatrix}$

The probability that it will rain on 3rd day is, 0.376
The probability that it will not rain on 3rd day is, 0.688 .....

A person owning a scooter has the opinion to switch over to a scooter, bike or a car next time with the probability of (0.3, 0.5, 0.2). If the transition probability matrix is, $\begin{bmatrix} \mathbf{0.4} & \mathbf{0.3} & \mathbf{0.3} \\ \mathbf{0.2} & \mathbf{0.5} & \mathbf{0.3} \\ \mathbf{0.25} & \mathbf{0.225} & \mathbf{0.5} \end{bmatrix}$.
What are the probabilities of vehicles to his fourth purchase?

We have P = $\begin{bmatrix} \mathbf{0.4} & \mathbf{0.3} & \mathbf{0.3} \\ \mathbf{0.2} & \mathbf{0.5} & \mathbf{0.3} \\ \mathbf{0.25} & \mathbf{0.225} & \mathbf{0.5} \end{bmatrix}$

P(1st purchase) = P(1) = P^1 = $\begin{bmatrix} \mathbf{0.4} & \mathbf{0.3} & \mathbf{0.3} \\ \mathbf{0.2} & \mathbf{0.5} & \mathbf{0.3} \\ \mathbf{0.25} & \mathbf{0.225} & \mathbf{0.5} \end{bmatrix}$

P(2nd purchase) = P(2) = P^2 = P. P = $\begin{bmatrix} \mathbf{0.295} & \mathbf{0.345} & \mathbf{0.36} \\ \mathbf{0.255} & \mathbf{0.358} & \mathbf{0.36} \\ \mathbf{0.275} & \mathbf{0.325} & \mathbf{0.4} \end{bmatrix}$

P(3rd purchase) = P(3) = P^3 = P. P. P = $\begin{bmatrix} \mathbf{0.277} & \mathbf{0.351} & \mathbf{0.372} \\ \mathbf{0.269} & \mathbf{0.359} & \mathbf{0.372} \\ \mathbf{0.275} & \mathbf{0.345} & \mathbf{0.380} \end{bmatrix}$

Hence, the probabilities of 4th purchase = (0.3, 0.5, 0.2) $\begin{bmatrix} \mathbf{0.277} & \mathbf{0.351} & \mathbf{0.372} \\ \mathbf{0.269} & \mathbf{0.359} & \mathbf{0.372} \\ \mathbf{0.275} & \mathbf{0.345} & \mathbf{0.380} \end{bmatrix}$

= (0.2726, 0.3538, 0.3736)

Three boys A, B and C are throwing a ball to each other. A is always throws the ball to B and B is always throws the ball to C, but C is just likely to throw the ball to B as to A. Show that the process is Markovian. Find the transition matrix and classify the states. Do all the states are ergodic?
If C was the 1st person to throw the ball, find the probabilities that a) A has the ball, b) B has the ball, c) C has the ball after three throws.

Probability of A is always throws the ball to B is, 1
Probability of B is always throws the ball to C is, 1
Probabilities of C is throws the ball to A and B are, 0.5 and 0.5
Probabilities of remaining all cases are, 0's

The transition probability matrix is, P = $\begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{0.5} & \mathbf{0.5} & \mathbf{0} \end{bmatrix}$

The states of row are depends only on states of column, but not other states or earlier states.

Hence, the process is Markovian or Markov chain.

Now,

P^1 = $\begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{0.5} & \mathbf{0.5} & \mathbf{0} \end{bmatrix}$

P^2 = $\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{0.5} & \mathbf{0.5} & \mathbf{0} \\ \mathbf{0} & \mathbf{0.5} & \mathbf{0.5} \end{bmatrix}$

P^3 = $\begin{bmatrix} \mathbf{0.5} & \mathbf{0.5} & \mathbf{0} \\ \mathbf{0} & \mathbf{0.5} & \mathbf{0.5} \\ \mathbf{0.25} & \mathbf{0.25} & \mathbf{0.5} \end{bmatrix}$

$$P^4 = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \\ 0.5 & 0.5 & 0 \end{bmatrix}$$

$$P^5 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.5 & 0.5 & 0 \end{bmatrix}$$

$$P^6 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.5 & 0.5 & 0 \end{bmatrix}, \ldots\ldots$$

the chain is irreducible.

Since, the chain is finite and irreducible, all the states are non-persistent.

The states are egodic.

Now,

If the ball was 1st thrown by C, the initial probability distribution is given by the row matrix, $A^{(0)} = [P_1^{(0)} \; P_2^{(0)} \; P_3^{(0)}] = [0 \; 0 \; 1]$

After three throws, the probability is given by

$$[P_1^{(0)} \; P_2^{(0)} \; P(0)_3] = [A^{(0)} \; P^3]$$

$$= [0 \; 0 \; 1] \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$$

$$= [0.25 \quad 0.25 \quad 0.25]$$

Thus, after three throws, the probability that the ball is with A is 0.25

B is 0.25

C is 0.25

Find the nature of the states of the Markov chain with t.p.m, $P = \begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix}$

We have, $$P = \begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix} = \mathbf{P}$$

$$P^4 = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} = P^2$$

$$P^5 = \begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix} = \mathbf{P}$$

$$P^6 = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} = P^2 \; \ldots.$$

Hence, the Markov chain is irreducible.

Since, Markov chain is irreducible and finite, all the states are non-null persistent.

All states are not ergodic.

In a cascade of binary communication channels, the symbols 1 and 0 are transmitted in successive stages. In any stage, the probability that a transmitted 1 is received as a 1 is 0.75 and the probability that 0 is received as a 0.5. Find the probabilities that i) 1 transmitted in the 1st stage is received correctly. ii) 0 transmitted in the 1st stage is received as 1 after the 3rd stage.

We have,

Probability that a transmitted 1 is received as 1 = 0.75

Probability that a transmitted 1 is received as 0 = 0.25

Probability that a transmitted 0 is received as 1 = 0.5

Probability that a transmitted 0 is received as 0 = 0.5

The transition matrix is $P = \begin{bmatrix} 0.75 & 0.25 \\ 0.5 & 0.5 \end{bmatrix}$

$$P^2 = \begin{bmatrix} 0.6875 & 0.3125 \\ 0.625 & 0.375 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 0.6719 & 0.3281 \\ 0.6563 & 0.3437 \end{bmatrix}$$

Therefore,

i) The probability that 1 is transmitted in the 1st stage and received correctly as 1 after the 3rd states is 0.6719

ii) The probability that 0 is transmitted in the 1st stage and received as 1 after the 3rd states is 0.6563

Check whether the following Markov chains are regular and ergodic?

i) $$\begin{bmatrix} 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 0.5 \end{bmatrix}$$

First we check the regularity of P(n) matrices for different n.

We observe that, when n is odd, P(n) is original matrix

When n is even, $P(n) = \begin{bmatrix} 0.5 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0.5 \end{bmatrix}$

For any value of n, we get matrices with zero only.

Therefore, the chain is not regular.

It is possible to go from state 0 to state 1 or state 2 (or) from state 1 to state 0 or state 3 and from state 2 to state 0 or state 1.

Hence, all the states are ergodic.

Now,

$$P^2 = \begin{bmatrix} 1.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.25 & 0.25 & 0.75 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 2 & 1 & 1 & 0.75 \\ 1 & 0.375 & 0.375 & 0.75 \\ 1 & 0.375 & 0.375 & 0.75 \\ 0.75 & 0.5 & 0.5 & 0.625 \end{bmatrix} \ldots.$$

Here, the matrices P^2, P^2, ... without zero elements.
We can also prove that P^m (m=2, 3, . . .)has all the entries with possible non-zero values.

Therefore, P is a regular matrix.

ii) $$\begin{bmatrix} 1 & 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 0.5 \end{bmatrix}$$

The transition probability matrix of a Markov chain $\{X_n\}$;n=1, 2, 3, . . .having three states 1, 2 and 3 is $P = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$ and the initial distribution is $P^{(0)} = (0.7, 0.2, 0.1)$

Find (i) $P\{X_2=3\}$, (ii) $P\{X_3=2, X_2=3, X_1=3, X_0=2\}$, (iii) $P\{X_3=2, X_2=3, X_1=3, X_0=3\}$

Given that, $P(X_0=1) = 0.7$, $P(X_0=2) = 0.2$, $P(X_0=3) = 0.1$

To find $P\{X_2=3\}$:

$$P = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$$

$$P^{(2)} = P^2 = \begin{bmatrix} 0.43 & 0.31 & 0.26 \\ 0.24 & 0.42 & 0.34 \\ 0.36 & 0.35 & 0.29 \end{bmatrix}$$

$$P\{X_2=3\} = \sum_{i=1}^{3} P\{X_2=3 / X_i=i\}.P\{X_0=i\}$$

$$= P\{X_2=3/X_0=1\}.P\{X_0=1\} + P\{X_2=3/X_0=2\}.P\{X_0=2\} + P\{X_2=3/X_0=3\}.P\{X_0=3\}$$

$$= P_{13}^{(2)} . P\{X_0=1\} + P_{23}^{(2)} . P\{X_0=2\} + P_{33}^{(2)} . P\{X_0=3\}$$
$$= 0.26\,(0.7) + 0.34\,(0.2) + 0.29\,(0.1)$$
$$= 0.279$$

$\underline{P\{X_2=3\} = 0.279}$

To find $P\{X_3=2, X_2=3, X_1=3, X_0=2\}$:

Consider, $P\{X_1=3, X_0=2\} = P\{X_1=3 / X_0=2\}.P\{X_0=2\}$

$= P_{23} . P\{X_0=2\}$
$= 0.2\ (0.2)$
$= 0.04$

$\underline{P\{X_1=3, X_0=2\} = 0.04}$

Now, $P\{X_2=3, X_1=3, X_0=2\} = P\{X_2=3 / X_1=3, X_0=2\} . P\{X_1=3, X_0=2\}$

$= P_{33} . P\{X_1=3, X_0=2\}$
$= 0.3\ (0.04)$
$= 0.012$

$\underline{P\{X_2=3, X_1=3, X_0=2\} = 0.012}$

Now, $P\{X_3=2, X_2=3, X_1=3, X_0=2\} = P\{X_3=2 / X_2=3, X_1=3, X_0=2\}. P\{X_2=3, X_1=3, X_0=2\}$

$= P_{32} . P\{X_2=3, X_1=3, X_0=2\}$
$= 0.4\ (0.012)$
$= 0.0048$

$\underline{P\{X_3=2, X_2=3, X_1=3, X_0=2\} = 0.0048}$

Therefore, $P\{X_3=2, X_2=3, X_1=3, X_0=2\} = 0.0048$

Also find $P\{X_3=2, X_2=3, X_1=3, X_0=3\} = ?$

1. A fair die thrown repeatedly. If X_n denotes the maximum of the numbers occurring in the first n throws. i) Find the tpm P of the Markov chain $\{X_n\}$, ii) P^2 & iii) $P\{X_2 = 6\}$

<u>Sol:</u> State space = {1, 2, 3, 4, 5, 6}

The tpm is formed using the following analysis,

Let X_n = Maximum of the numbers occurring in the first n throws = 3 (Say)

Then, $X_n = 3$; if $(n+1)^{th}$ throw results is 1 or 2 or 3

$= 4$; if $(n+1)^{th}$ throw results is 4

$= 5$; if $(n+1)^{th}$ throw results is 5

$= 6$; if $(n+1)^{th}$ throw results is 6

Therefore, $\mathbf{P\{X_{n+1} = 3 / X_n = 3\} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6}}$

$\mathbf{P\{X_{n+1} = i / X_n = 3\} = \frac{1}{6}}$ **when i = 4, 5, 6**

i) The transition probability matrix (t.p.m) of the Markov chain is,

$$P = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{2}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & \frac{3}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & \frac{4}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & \frac{5}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & \frac{6}{6} \end{bmatrix}$$

ii) $$\mathbf{P^2} = \frac{1}{36}\begin{bmatrix} 1 & 3 & 5 & 7 & 9 & 11 \\ 0 & 4 & 5 & 7 & 9 & 11 \\ 0 & 0 & 9 & 7 & 9 & 11 \\ 0 & 0 & 0 & 16 & 9 & 11 \\ 0 & 0 & 0 & 0 & 25 & 11 \\ 0 & 0 & 0 & 0 & 0 & 36 \end{bmatrix}$$

Since, all the values of 1, 2, 3, 4, 5, 6 are equally likely. So, the initial state probability distribution is, $\mathbf{P^{(0)}} = \left[\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right]$

iii) $$\mathbf{P\{X_2 = 6\} = \sum_{i=1}^{6} P\{X_2 = 6 / X_0 = i\} . P\{X_0 = i\}}$$

$$= \frac{1}{6}\sum_{i=1}^{6} P_{i6}^{(2)}$$

$$= \frac{1}{6}\left[\frac{11}{36} + \frac{11}{36} + \frac{11}{36} + \frac{11}{36} + 6\right]$$

$$= \frac{1}{6(36)}[11 + 11 + 11 + 11 + 11 + 36]$$

- $\mathbf{P\{X_2 = 6\} = \frac{91}{216} = 0.4213}$

2. Three boys A, B, C are throwing a ball to each other. A is always throws the ball to B, B always throws the ball to C but C just as likely to throw the ball to B as A. show that the process is Markovian. Find the transition matrix. Classify the states. Do all the states are ergodic?

Sol:

The transition probability matrix of the process $\{X_n\}$ is, $\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.5 & 0.5 & 0 \end{bmatrix}$

State $\{X_n\}$ denotes only on states of $\{X_{n-1}\}$ but not on states of $X_{n-1}, X_{n-1}, \ldots$ or earlier states.

Therefore, $\{X_n\}$ is a Markov chain.

Now, $\mathbf{P}^2 = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{0.5} & \mathbf{0.5} & \mathbf{0} \\ \mathbf{0} & \mathbf{0.5} & \mathbf{0.5} \end{bmatrix}$ $\mathbf{P}^3 = \begin{bmatrix} \mathbf{0.5} & \mathbf{0.5} & \mathbf{0} \\ \mathbf{0} & \mathbf{0.5} & \mathbf{0.5} \\ \mathbf{0.25} & \mathbf{0.25} & \mathbf{0.5} \end{bmatrix}$

$$\mathbf{P}_{11}^{(3)} > 0, \mathbf{P}_{13}^{(2)} > 0, \mathbf{P}_{21}^{(2)} > 0, \mathbf{P}_{22}^{(2)} > 0, \mathbf{P}_{33}^{(2)} > 0$$

Therefore, the chain is irreducible.

$\mathbf{P}^4 = \begin{bmatrix} \mathbf{0} & \mathbf{0.5} & \mathbf{0.5} \\ \mathbf{0.25} & \mathbf{0.25} & \mathbf{0.5} \\ \mathbf{0.25} & \mathbf{0.5} & \mathbf{0.25} \end{bmatrix}$ $\mathbf{P}^5 = \begin{bmatrix} \mathbf{0.25} & \mathbf{0.25} & \mathbf{0.5} \\ \mathbf{0.25} & \mathbf{0.5} & \mathbf{0.25} \\ \mathbf{0.125} & \mathbf{0.375} & \mathbf{0.25} \end{bmatrix}$ $\mathbf{P}^6 = \begin{bmatrix} \mathbf{0.25} & \mathbf{0.5} & \mathbf{0.25} \\ \mathbf{0.25} & \mathbf{0.375} & \mathbf{0.5} \\ \mathbf{0.125} & \mathbf{0.375} & \mathbf{0.375} \end{bmatrix}$ ... .

G. C. D of 2, 3, 4, 5, 6 . . . . = 1

The state's 2 & 3 (i.e., B & C) are periodic with period 1, i.e., a periodic.

We note that $\mathbf{P}_{ii}^{(3)}, \mathbf{P}_{ii}^{(5)}, \mathbf{P}_{ii}^{(6)}$ ... **are > 0** and G. C. D of 3, 5, 6 , ... =1

Therefore the state 1 (i.e., state A) is periodic with period 1, i.e., a periodic.

Since the chain is finite and irreducible, all its states are non-null persistent.

Hence, all the states are **ergodic**.

3. The tpm of a Markov chain is given by $\begin{bmatrix} \mathbf{0.3} & \mathbf{0.7} & \mathbf{0} \\ \mathbf{0.1} & \mathbf{0.4} & \mathbf{0.5} \\ \mathbf{0} & \mathbf{0.2} & \mathbf{0.8} \end{bmatrix}$ irreducible?

Sol: In this matrix, the chain starts from state 0 to state 1 with probability 0.7, from state 1 to state 2 with probability 0.5

Thus, it is possible to go from state 0 to state 2

Therefore, the chain is irreducible & the states are recurrent.

4. Is the tpm $\begin{bmatrix} \mathbf{0.4} & \mathbf{0.6} & \mathbf{0} & \mathbf{0} \\ \mathbf{0.3} & \mathbf{0.7} & \mathbf{0} & \mathbf{0} \\ \mathbf{0.2} & \mathbf{0.4} & \mathbf{0.1} & \mathbf{0.3} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}$ irreducible?

Sol: Given tpm is, $\mathbf{P} = \begin{bmatrix} \mathbf{0.4} & \mathbf{0.6} & \mathbf{0} & \mathbf{0} \\ \mathbf{0.3} & \mathbf{0.7} & \mathbf{0} & \mathbf{0} \\ \mathbf{0.2} & \mathbf{0.4} & \mathbf{0.1} & \mathbf{0.3} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}$, $\mathbf{P}^2 = \begin{bmatrix} \mathbf{0.34} & \mathbf{0.66} & \mathbf{0} & \mathbf{0} \\ \mathbf{0.33} & \mathbf{0.69} & \mathbf{0} & \mathbf{0} \\ \mathbf{0.22} & \mathbf{0.44} & \mathbf{0.01} & \mathbf{0.33} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}$,

$\mathbf{P}^3 = \begin{bmatrix} \mathbf{0.334} & \mathbf{0.666} & \mathbf{0} & \mathbf{0} \\ \mathbf{0.333} & \mathbf{0.667} & \mathbf{0} & \mathbf{0} \\ \mathbf{0.222} & \mathbf{0.444} & \mathbf{0.001} & \mathbf{0.003} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}$

w. k. t, to prove, a matrix is irreducible that the sum of each row of the matrix must be equal to 1 but in P^3, the sum of second row is not equal to 1

Hence, the given matrix is not irreducible.

5. If the tpm of a Markov chain is $\begin{bmatrix} 0 & 1 \\ 0.5 & 0.5 \end{bmatrix}$ find the steady state distribution of the chain.

Sol: Given $P = \begin{bmatrix} 0 & 1 \\ 0.5 & 0.5 \end{bmatrix}$

Let $X = [x_1 \quad x_2]$ is the steady-state distribution of the chain

Then X=XP and $x_1 + x_2 = 1$ → Equ 1

$[x_1 \quad x_2] = [x_1 \quad x_2] * \begin{bmatrix} 0 & 1 \\ 0.5 & 0.5 \end{bmatrix}$

$[x_1 \quad x_2] = \left[\frac{x_2}{2} \quad x_1 + \frac{x_2}{2}\right]$

$x_1 = \frac{x_2}{2}$ => $2x_1$=x_2 → Equ 2

$x_2 = x_1 + \frac{x_2}{2}$ → Equ 3

Substitute Equ 2 in Equ 1, we get $x_1 = \frac{1}{3}$

From Equ 2, $x_2 = \frac{2}{3}$

Hence, $\left[\frac{1}{3} \quad \frac{2}{3}\right]$ be the steady-state distribution of the Markov chain.

6. Study of a passage of English text to find a vowel followed by a vowel or a consonant followed by a consonant or a vowel reveal the following transition probability matrix $\begin{bmatrix} 0.12 & 0.88 \\ 0.54 & 0.46 \end{bmatrix}$

Find the percentage of the letters in the text book which are expected to be the vowels.

Sol: Given $P = \begin{bmatrix} 0.12 & 0.88 \\ 0.54 & 0.46 \end{bmatrix}$

To find the limiting probabilities, we have X=XP and $x_1 + x_2 = 1$ → Equ 1

$$[x_1 \quad x_2] = [x_1 \quad x_2] * \begin{bmatrix} 0.12 & 0.88 \\ 0.54 & 0.46 \end{bmatrix}$$

$x_1 = 0.12x_1 + 0.54x_2$ → Equ 2

$x_2 = 0.88x_1 + 0.46x_2$ → Equ 3

On substituting from Equ 1, $x_2 = 1 - x_1$ in Equ 2

We get, $x_1 = 0.3803$

$x_2 = 1 - x_1 = 1 - 0.3803$

$x_2 = 0.6197$

Hence, the percentage of letters in the text book which are expected to be vowels is 38%

7. The school of international studies for population found out by its survey that the mobility of the population of a state to village, town and city is the following percentage.

from \ To	Village	Town	City
Village	30%	20%	50%
Town	30%	50%	20%
City	10%	40%	50%

Present population of village, town and city are 0.4, 0.3 and 0.3 respectively.

What will the proportion of population in village, town and city after two years?

Find the proportions in the long run.

Sol:

Given that, the school of international studies for population found out by its survey that the mobility of the population of a state to village, town and city are,

from \ To	Village	Town	City
Village	30%	20%	50%
Town	30%	50%	20%
City	10%	40%	50%

This can be written as, $P = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$

The present population proportion, $P_o = [0.4 \quad 0.3 \quad 0.3]$

The proportion of population in village, town and city after one year, $P_1 = P_0 * P$

$$\Rightarrow \quad P_1 = [0.4 \quad 0.3 \quad 0.3] * \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{bmatrix} = [0.24 \quad 0.35 \quad 0.4]$$

Hence, the proportion of population in village, town and city after one year, $P_1 = [0.24 \quad 0.35 \quad 0.4]$

The proportion of population in village, town and city after two years, $P_2 = P_1 * P$

$$\Rightarrow \quad P_2 = [0.24 \quad 0.35 \quad 0.4] * \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{bmatrix} = [0.218 \quad 0.387 \quad 0.395]$$

Hence, the proportion of population in village, town & city after two years, $P_2 = [0.218 \quad 0.387 \quad 0.395]$

Hence, the proportions of population in village, town & city after two years are, 0.218, 0.387 & 0.395

- The proportions of population in village, town & city in long run

i.e., the steady state distribution of the chain, then we have X=XP and $x_1 + x_2 + x_3 = 1$

$$[x_1 \quad x_2 \quad x_3] = [x_1 \quad x_2 \quad x_3] * \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

$We\ get, x_1 = 0.3x_1 + 0.3x_2 + 0.1x_3$

⇨ $0.7x_1 = 0.3x_2 + 0.1x_3$ → Equ 1

Also, we get, $x_2 = 0.2x_1 + 0.5x_2 + 0.4x_3$

⇨ $0.5x_2 = 0.2x_1 + 0.4x_3$ → Equ 2

And also, we get, $x_3 = 0.5x_1 + 0.2x_2 + 0.5x_3$

⇨ $0.5x_3 = 0.5x_1 + 0.2x_2$ → Equ 3

Now, $x_1 + x_2 + x_3 = 1$ => $x_1 = 1 - x_2 - x_3$ → Equ 4

On multiplying Equ 4 with 0.7 and substituting in Equ 1

We get, $x_2 + 0.8x_3 = 0.7$ → Equ 5

On multiplying Equ 2 with 0.5 and Equ 3 with 0.2, then we get $x_2 = 1.034x_3$ → Equ 6

Now, on substituting Equ 6 (x_2) in Equ 5, we get $x_3 = 0.382$

Now, on substituting $x_3 = 0.382$ in Equ 6, we get $x_2 = 0.394$

Now, on substituting $x_3 = 0.382$ $x_2 = 0.394$ in Equ 4, we get $x_1 = 0.224$

Therefore, $X = [x_1 \quad x_2 \quad x_3] = [0.224 \quad 0.394 \quad 0.382]$

Hence, the long run proportion of the population state to village, town and city is, 22.4%, 39.4% and 38.2% respectively.

8. If the transition probability matrix of market shares of three brands A, B and C is $\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.8 & 0.1 & 0.1 \\ 0.35 & 0.25 & 0.4 \end{bmatrix}$ and the initial market shares are 50%, 25% and 25%.

Find the market shares in second and third periods and the limiting probabilities.

Sol:

Given that, the initial market shares of three brands A, B and C are 50%, 25% and 25% which can be written as $P_0 = [0.5 \quad 0.25 \quad 0.25]$

The tpm is, $P = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.8 & 0.1 & 0.1 \\ 0.35 & 0.25 & 0.4 \end{bmatrix}$

$P^2 = \begin{bmatrix} 0.5050 & 0.225 & 0.27 \\ 0.435 & 0.275 & 0.29 \\ 0.48 & 0.23 & 0.29 \end{bmatrix}$ and $P^3 = \begin{bmatrix} 0.4145 & 0.2085 & 0.19 \\ 0.4675 & 0.2105 & 0.1985 \\ 0.4652 & 0.2307 & 0.221 \end{bmatrix}$

The market share in the second period = $P_0 * P^2$

$$[0.5 \quad 0.25 \quad 0.25] * \begin{bmatrix} 0.5050 & 0.225 & 0.27 \\ 0.435 & 0.275 & 0.29 \\ 0.48 & 0.23 & 0.29 \end{bmatrix} = [0.4813 \quad 0.2387 \quad 0.2800]$$

The market share in the third period = $P_0 * P^3$

$$[0.5 \quad 0.25 \quad 0.25] * \begin{bmatrix} 0.4145 & 0.2085 & 0.19 \\ 0.4675 & 0.2105 & 0.1985 \\ 0.4652 & 0.2307 & 0.221 \end{bmatrix} = [0.4815 \quad 0.2382 \quad 0.2802]$$

Hence, the market shares in the third period are, 48.1%, 23.8% and 28.02%.

- The limiting probability can be interpreted as the long run market shares that can be reached to an equilibrium

$X = [x_1 \quad x_2 \quad x_3]$ then,

X=XP and $x_1 + x_2 + x_3 = 1$ → Equ 1

$$[x_1 \quad x_2 \quad x_3] = [x_1 \quad x_2 \quad x_3] * \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.8 & 0.1 & 0.1 \\ 0.35 & 0.25 & 0.4 \end{bmatrix}$$

We get, $x_1 = 0.4x_1 + 0.8x_2 + 0.35x_3$ => $0.6x_1 = 0.8x_2 + 0.35x_3$ → Equ 2

$x_2 = 0.3x_1 + 0.1x_2 + 0.25x_3$ => $0.9x_2 = 0.3x_1 + 0.25x_3$ → Equ 3

$x_3 = 0.3x_1 + 0.1x_2 + 0.4x_3$ => $0.6\,x_3 = 0.3x_1 + 0.1x_2$ → Equ 4

On solving the above equations, we get,

$x_1 = 0.4813,\ x_2 = 0.2383\ and\ x_3 = 0.28.4$

9. If the transition probability matrix of market shares of three brands A, B and C is $\begin{bmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.5 & 0.5 \\ 0.6 & 0.4 & 0 \end{bmatrix}$ and the initial market shares are 20%, 30% and 50%.

Find the market shares in second and third periods and the limiting probabilities.