MA303BS: COMPUTER ORIENTED STATISTICAL METHODS

B.TECH II Year I Sem.

L	T	P	C
3	1	0	4

Pre-requisites: Mathematics courses of first year of study.

Course Objectives: To learn

- The theory of Probability, and probability distributions of single and multiple random variables
- The sampling theory and testing of hypothesis and making inferences
- Stochastic process and Markov chains.

Course Outcomes: After learning the contents of this paper the student must be able to

- Apply the concepts of probability and distributions to some case studies
- Correlate the material of one unit to the material in other units
- Resolve the potential misconceptions and hazards in each topic of study.

UNIT - I

Probability: Sample Space, Events, Counting Sample Points, Probability of an Event, Additive Rules, Conditional Probability, Independence, and the Product Rule, Bayes' Rule.

Random Variables and Probability Distributions: Concept of a Random Variable, Discrete Probability Distributions, Continuous Probability Distributions, Statistical Independence.

UNIT - II

Mathematical Expectation: Mean of a Random Variable, Variance and Covariance of Random Variables, Means and Variances of Linear Combinations of Random Variables, Chebyshev's Theorem.

Discrete Probability Distributions: Introduction and Motivation, Binomial, Distribution, Geometric Distributions and Poisson distribution.

UNIT - III

Continuous Probability Distributions : Continuous Uniform Distribution, Normal Distribution, Areas under the Normal Curve, Applications of the Normal Distribution, Normal Approximation to the Binomial, Gamma and Exponential Distributions.

Fundamental Sampling Distributions: Random Sampling, Some Important Statistics, Sampling Distributions, Sampling Distribution of Means and the Central Limit Theorem, Sampling Distribution of S2, t –Distribution, F-Distribution.

UNIT - IV

Estimation & Tests of Hypotheses: Introduction, Statistical Inference, Classical Methods of Estimation.: Estimating the Mean, Standard Error of a Point Estimate, Prediction Intervals, Tolerance Limits, Estimating the Variance, Estimating a Proportion for single mean , Difference between Two Means, between Two Proportions for Two Samples and Maximum Likelihood Estimation.

Statistical Hypotheses: General Concepts, Testing a Statistical Hypothesis, Tests Concerning a Single Mean, Tests on Two Means, Test on a Single Proportion, Two Samples: Tests on Two Proportions.

UNIT - V

Stochastic Processes and Markov Chains: Introduction to Stochastic processes- Markov process. Transition Probability, Transition Probability Matrix, First order and Higher order Markov process, n-step transition probabilities, Markov chain, Steady state condition, Markov analysis.

TEXT BOOKS:

1. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, Probability & Statistics for Engineers & Scientists, 9th Ed. Pearson Publishers.
2. S C Gupta and V K Kapoor, Fundamentals of Mathematical statistics, Khanna publications.
3. S. D. Sharma, Operations Research, Kedarnath and Ramnath Publishers, Meerut, Delhi

REFERENCE BOOKS:

1. T.T. Soong, Fundamentals of Probability and Statistics for Engineers, John Wiley & Sons Ltd, 2004.
2. Sheldon M Ross, Probability and statistics for Engineers and scientists, Academic Press.

UNIT-I

PROBABILITY AND RANDOM VARIABLES

Probability

Trial and Event: Consider an experiment, which though repeated under essential and identical conditions, does not give a unique result but may result in any one of the several possible outcomes. The experiment is known as **Trial** and the outcome is called **Event**

E.g. (1) Throwing a dice experiment getting the no's 1,2,3,4,5,6 (event)

(2) Tossing a coin experiment and getting head or tail (event)

Exhaustive Events:

The total no. of possible outcomes in any trial is called exhaustive

event. E.g.: (1) In tossing of a coin experiment there are two

exhaustive events.

(2) In throwing an n-dice experiment, there are 6^n exhaustive events.

Favorable event:

The no of cases favorable to an event in a trial is the no of outcomes which entities the happening of the event.

E.g. (1) In tossing a coin, there is one and only one favorable case to get either head or tail.

Mutually exclusive Event: If two or more of them cannot happen simultaneously in the same trial then the event are called mutually exclusive event.

E.g. In throwing a dice experiment, the events 1,2,3, 6 are M.E. events

Equally likely Events: Outcomes of events are said to be equally likely if there is no reason for one to be preferred over other. E.g. tossing a coin. Chance of getting 1,2,3,4,5,6 is equally likely.

Independent Event:

Several events are said to be independent if the happening or the non-happening of the event is not affected by the concerning of the occurrence of any one of the remaining events.

An event that always happen is called **Certain event,** it is denoted by

'S'. An event that never happens is called **Impossible event**, it is

denoted by '□'.

Eg: In tossing a coin and throwing a die, getting head or tail is independent of getting no's 1 or 2 or 3 or 4 or 5 or 6.

Definition: probability (Mathematical Definition)

If a trial results in n-exhaustive mutually exclusive, and equally likely cases and m of them are favorable to the happening of an event E then the probability of an event E is denoted by P(E) and is defined as

$$P(E) = \frac{\text{no of favourable cases to event}}{\text{Total noof exaustivecases}} = \frac{m}{n}$$

Sample Space:

The set of all possible outcomes of a random experiment is called Sample Space .The elements of this set are called sample points. Sample Space is denoted by S.

Eg. (1) In throwing two dies experiment, Sample S contains 36 Sample

points. S = {(1,1) ,(1,2) ,----------(1,6), --------(6,1),(6,2), (6,6)}

Eg. (2) In tossing two coins experiment , S = {HH ,HT,TH,TT}

A sample space is called **discrete** if it contains only finitely or infinitely many points which can be arranged into a simple sequence $w_1, w_2, \ldots$ while a sample space containing non denumerable no. of
points is called a continuous sample space.

Statistical or Empirical Probability:

If a trial is repeated a no. of times under essential homogenous and identical conditions, then the limiting value of the ratio of the no. of times the event happens to the total no. of trials, as the number of trials become indefinitely large, is called the probability of happening of the event.(It is assumed the limit is finite and unique)

Symbolically, if in 'n' trials and events E happens 'm' times , then the probability 'p' of the happening of E is given by $p = P(E) = \lim_{n \to \infty} \frac{m}{n}$.

An event E is called **elementary event** if it consists only one element. An event, which is not elementary, is called **compound event.**

Conditional Probability:

The ***conditional probability*** of an event *B* is the probability that the event will occur given the knowledge that an event *A* has already occurred. This probability is written *P(B|A)*, notation for the *probability of B given A*. In the case where events *A* and *B* are *independent* (where event *A* has no effect on the probability of event *B*), the conditional probability of event *B* given event *A* is simply the probability of event *B*, that is *P(B)*.

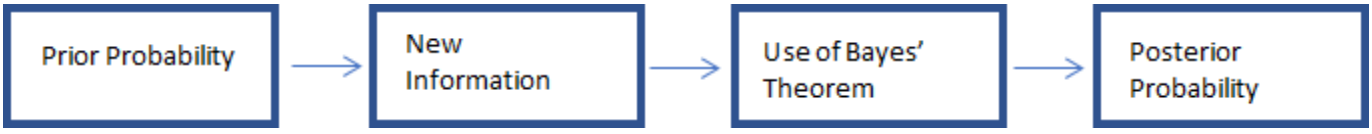
If events *A* and *B* are not independent, then the probability of the *intersection of A and B* (the probability that both events occur) is defined by
P(A and B) = P(A)P(B|A).

From this definition, the conditional probability *P(B|A)* is easily obtained by dividing by *P(A)*:

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

Baye's Theorem

We are quite familiar with probability and its calculation. From one known probability we can go on calculating others. But can we use all the prior information to calculate or to measure the chance of some events happened in past? This is the posterior probability. Bayes theorem calculates the posterior probability of a new event using a prior probability of some events.



Baye's theorem, sometimes, also calculates the probability of some future events.

Theorem

If $E_1, E_2, E_3, \dots, E_n$ are mutually disjoint events with $P(E_i) \neq 0$, (i = 1, 2, …, n), then for any arbitrary event A which is a subset of the union of events E_i such that P(A) > 0, we have

$P(E_i \mid A) = [P(E_i).P(A \mid E_i)] \div [\sum_i P(E_i).P(A \mid E_i)] = [P(E_i).P(A \mid E_i)] \div P(A)$, $E_1, E_2, E_3, \dots,$ E_n represents the partition of the sample space S

Proof

A is a subset of the union of E_i, i.e., $A \subset \cup E_i$, we have, $A = A \cap (\cup E_i) = \cup(A \cap E_i)$. Also, $(A \cap E_i)$ subset E_i, (i = 1, 2, …, n) are mutually disjoint events as E_i's are mutually disjoint. So, $P(A) = P[\cup(A \cap E_i)] = \sum_i P(A \cap E_i) = \sum_i P(E_i).\ P(A \mid E_i)$.

Also, $P(A \cap E_i) = P(A).\ P(E_i \mid A)$

or, $P(E_i \mid A) = P(A \cap E_i) / P(A) = [P(E_i).\ P(A \mid E_i)] \div [\sum_i P(E_i).\ P(A \mid E_i)]$.

problems

1. Let us consider the situation where a child has three bags offruits in which Bag 1 has 5 apples and 3 oranges, Bag 2 has 3 apples and 6 oranges and Bag 3 has 2 apples and 3 oranges. One fruit is drawn at random from one of the bags. It was an apple. Let us calculate the probability that the chosen fruit was apple and was drawn from Bag 3.

Here, we can calculate the probability of selecting the bags, $P(E_1) = P(E_2) = P(E_3) = 1/3$. The probability of drawing out of apple from Bag 1, $P(A \mid E_1) = 5/8$, from Bag 2, $P(A|E_2) = 3/9 = 1/3$, from Bag 3, $P(A \mid E_3) = 2/5$.

We have to calculate the probability of drawing a fruit given that we have chosen the Bag 3. The probability of drawing a fruit from Bag 3 given that the chosen fruit is an apple is $P(E_3|A)$. The Baye's formula helps us to calculate the probability, which is

$P(E_3| A) = [P(E_3).\ P(A \mid E_3)] \div [P(E_1) \times P(A \mid E_1) + P(E_2) \times P(A \mid E_2) + P(E_3) \times P(A|E_3)]$

$\Rightarrow P(E_3| A)\ 0 = 1/3 \times 2/5 \div [(1/3 \times 5/8) + (1/3 \times 3/9) + (1/3 \times 2/5)] = (2/15) \div (163/360) = 48/163$.

2. One box is chosen at random and three balls are drawn from it. They all are of different colors. What is the probability that they come from boxes I, II or III?

Solution: Let A be the event of drawing three balls. E_1, E_2, E_3 represent the events of selecting Box I, II and III respectively.

$P(E_1) = P(E_2) = P(E_3) = 1/3$.

The probability of drawing three balls given that the Box I is

selected, $P(A|E_1) = ({}^1C_1 \times {}^2C_1 \times {}^3C_1) \div {}^6C_3 = 3/10$.

$P(A|E_2) = (1 \times 1 \times 2) \div {}^4C_3 = 1/2$.
$P(A|E_3) = (5 \times 4 \times 1) \div {}^{10}C_3 = 1/6$.
And $P(E_1)\times P(A|E_1) + P(E_2)\times P(A|E_2) + P(E_3)\times P(A|E_3) = 29/90$.

We can calculate the probabilities by using Baye's theorem. So, the required probability,

$P(E_1|A) = [P(E_1|A).\ P(A|E_1)] \div [P(E_1)\times P(A|E_1) + P(E_2)\times P(A|E_2) + P(E_3)\times P(A|E_3)] = (1/10) \div (29/90) = 9/29$.

$P(E_2|A) = [P(E_2|A).\ P(A|E_2)] \div [P(E_1)\times P(A|E_1) + P(E_2)\times P(A|E_2) + P(E_3)\times P(A|E_3)] = (1/6) \div (29/90) = 15/29$.

$P(E_3|A) = 1 - [P(E_1|A) + P(E_2|A)] = 1 - [(9/29) + (15/29)] = 1 - 24/29 = 5/29$.

Random Variable

A Random Variable X is a real valued function from sample space S to a real number R.

(or)

A Random Variable X is a real number which is determined by the outcomes of the random experiment.

Eg:1.Tosssing 2 coins simultaneously

Sample space ={HH,HT,TH,TT}

Let the random variable be getting number of heads then

X(S)={0,1,2}.

2.Sum of the two numbers on throwing 2 dice

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

Types of Random Variables:

1.**Discrete Random Variables :** A Random Variable X is said to be discrete if it takes only the values of the set {0,1,2…..n}.

Eg:1.Tosssing a coin, throwing a dice,number of defective items in a bag.

2.Continuous Random Variables: A Random Variable X which takes all possible values in a given interval of domain.

Eg: Heights, weights of students in a class.

Discrete Probability Distribution:

Let x is a Discrete Random Variable with possible outcomes $x_1, x_2, x_3 \ldots . x_n$ having probabilities $p(x_i) for\ i = 1,2 \ldots n$ .If $p(x_i) > 0\ and\ \sum_{i=1}^{n} p(x_i) = 1$ then the function $p(x_i)$ is called **Probability mass function** of a random variable X and $\{ x_i, p(x_i)\}\ for\ i = 1,2 \ldots n$ is called **Discrete Probability Distribution.**

Eg:Tosssing 2 coins simultaneously

Sample space ={HH,HT,TH,TT}

Let the random variable be getting number of heads then

X(S)={0,1,2}.

Probability of getting no heads = $\frac{1}{4}$, Probability of getting 1 head = $\frac{1}{2}$, Probability of getting 2 heads = $\frac{1}{4}$

∴ Discrete Probability Distribution is

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

Cummulative Distribution function is given by $F(x) = p[X \leq x] = \sum_{i=0}^{x} p(x_i)$.

Properties of Cummulative Distribution function:

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

2. $P[a \leq x \leq b] = F(b) - F(a) - P[X = a]$

3. $P[a < x \leq b] = F(b) - F(a)$

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

Mean: The mean of the discrete Probability Distribution is defined as

$\mu = \frac{\sum_{i=1}^{n} x_i p(x_i)}{\sum_{i=1}^{n} p(x_i)} = \sum_{i=1}^{n} x_i p(x_i)$ since $\sum_{i=1}^{n} p(x_i) = 1$

Expectation: : The Expectation of the discrete Probability Distribution is defined as

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

In general, $E(g(x)) = \sum_{i=1}^{n} g(x_i)p(x_i)$

Properties:

1) $E(X) = \mu$

2) $E(X) = k\,E(X)$

3) $E(X+k) = E(X) + k$

4)) $E(aX \pm b) = aE(X) \pm b$

Variance: The variance of the discrete Probability Distribution is defined as

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

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

$= \sum x_i{}^2 p_i - \mu^2$

Properties:

1) $V(c) = 0$ *where c is a constant*

2) $V(kX) = k^2\,V(X)$

3) $V(X+k) = V(X)$

4) $V(aX \pm b) = a^2\,V(X)$

PROBLEMS

1.If 3 cars are selected randomly from 6 cars having 2 defective cars.

a)Find the Probability distribution of defective cars.

b)Find the Expected number of defective cars.

Sol: Number of ways to select 3 cars from 6 cars = ${}^{6}c_{3}$

Let random variable X(S) = Number of defective cars = {0,1,2}

Probability of non defective cars = $\frac{{}^{4}c_{3}\,{}^{2}c_{0}}{{}^{6}c_{3}} = \frac{1}{5}$

Probability of one defective cars = $\frac{{}^{4}c_{2}\,{}^{2}c_{1}}{{}^{6}c_{3}} = \frac{3}{5}$

Probability of two defective cars = $\frac{{}^{4}c_{1}\,{}^{2}c_{2}}{{}^{6}c_{3}} = \frac{1}{5}$

Clearly , $p(x_i) > 0$ *and* $\sum_{i=1}^{n} p(x_i) = 1$

Probability distribution of defective cars is

x_i	0	1	2
$p(x_i)$	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{1}{5}$

Expected number of defective cars = $\sum_{i=1}^{n} x_i p(x_i) = 0\left(\frac{1}{5}\right) + 1\left(\frac{3}{5}\right) + 2\left(\frac{1}{5}\right) = 1$

2.Let X be a random variable of sum of two numbers in throwing two fair dice.Find the probability distribution of X,mean,variance.

Sol:Sample space of throwing two dices = S ={(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)
(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)
(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)
(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)
(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)
(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)}

$\therefore n(S) = 36.$

Let X = Sum of two numbers in throwing two dice = {2,3,4,5,6,7,8,9,10,11,12}

X	Favourable cases	No of Favourable cases	$p(x)$
2	(1,1)	1	$\frac{1}{36}$
3	(2,1),)(1,2)	2	$\frac{2}{36}$
4	(3,1),(2,2),(1,3)	3	$\frac{3}{36}$
5	(4,1),(3,2),(2,3),(1,4)	4	$\frac{4}{36}$
6	(5,1),(4,2),(3,3),(2,4),(1,5)	5	$\frac{5}{36}$
7	(6,1),(5,2),(4,3),(3,4),(2,5),(1,6)	6	$\frac{6}{36}$
8	(6,2),(5,3),(4,4),(3,5),(2,6)	5	$\frac{5}{36}$
9	(6,3),(5,4),(4,5),(3,6)	4	$\frac{4}{36}$

10	(6,4),(5,5),(4,6)	3	$\frac{2}{36}$
11	(6,5),(5,6)	2	$\frac{1}{36}$
12	(6,6)	1	

Clearly , $p(x_i) > 0 \; and \sum_{i=1}^{n} p(x_i) = 1$

Probability distribution is given by

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

$$Mean = \mu = \sum_{i=1}^{n} x_i p(x_i)$$

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

$$= 7.$$

$$Variance = V(X) = \sum x_i^{\,2} p_i - \mu^2$$

$$= 4\left(\frac{1}{36}\right) + 9\left(\frac{2}{36}\right) + 16\left(\frac{3}{36}\right) + 25\left(\frac{4}{36}\right) + 36\left(\frac{5}{36}\right) + 49\left(\frac{6}{36}\right) + 64\left(\frac{5}{36}\right) + 81\left(\frac{4}{36}\right) + 100\left(\frac{3}{36}\right) + 121\left(\frac{2}{36}\right) + 144\left(\frac{1}{36}\right) - 49$$

$$\therefore Variance = 5.83$$

3. .Let X be a random variable of maximum of two numbers in throwing two fair dice simultaneously.Find the

a)probability distribution of X b)mean c)variance d)P(1<x<4) e)P(2≤ $x \leq 4$).

Sol: Sample space of throwing two dices = S ={(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)

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

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

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

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

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

$\therefore n(S) = 36.$

Let X = Maximum of two numbers in throwing two dice = {1,2,3,4,5,6,}

X	Favourable cases	No of Favourable cases	$p(x)$
1	(1,1)	1	$\frac{1}{36}$
2	(2,1),)(1,2),(2,2)	3	$\frac{3}{36}$
3	(3,1),(1,3),(2,3)(3,3),(3,2)	5	$\frac{5}{36}$
4	(1,4),(4,1),(4,2),(2,4)(4,3),(3,4),(4,4)	7	$\frac{7}{36}$
5	(1,5),(5,1),(2,5),(5,2)(3,5),(5,3),(5,4),(4,5),(5,5)	9	$\frac{9}{36}$
6	(1,6)(6,1),(6,2),(2,6),(6,3),(3,6),(4,6),(6,4),(6,5)(5,6),(6,6)	11	$\frac{11}{36}$

Clearly , $p(x_i) > 0 \; and \sum_{i=1}^{n} p(x_i) = 1$

Probability distribution is given by

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

$$Mean = \mu = \sum_{i=1}^{n} x_i p(x_i) = 1\left(\frac{1}{36}\right) + 2\left(\frac{3}{36}\right) + 3\left(\frac{5}{36}\right) + 4\left(\frac{7}{36}\right) + 5\left(\frac{9}{36}\right) + 6\left(\frac{11}{36}\right)$$

= 4.4 7.

$Variance = V(X) = \sum x_i^2 p_i - \mu^2$

$= 1\left(\frac{1}{36}\right) + 4\left(\frac{3}{36}\right) + 9\left(\frac{5}{36}\right) + 16\left(\frac{7}{36}\right) + 25\left(\frac{9}{36}\right) + 36\left(\frac{11}{36}\right)$

$\therefore Variance = 1.99.$

4.A random variable X has the following probability function

x_i	-3	-2	-1	0	1	2	3
$p(x_i)$	k	0.1	k	0.2	2k	0.4	2k

Find k,mean,variance.

Sol: We know that $\sum_{i=1}^{n} p(x_i) = 1$

i.e k+0.1+k+0.2+2k+0.4+2k = 1

i.e 6k+0.7 = 1 $\therefore k = 0.05$

$Mean = \mu = \sum_{i=1}^{n} x_i p(x_i) = k(-3) + 0.1(-2) + k(-1) + 2k(1) + 2(0.4) + 3(2k)$

= 0.8.

$Variance = V(X) = \sum x_i^2 p_i - \mu^2$

$= k(-3)^2 + 0.1^2(-2) + k(-1)^2 + 2k(1) + 4(0.4) + 9(2k)$

$\therefore Variance = 2.86.$

Continuous Probability distribution:

Let X be a continuous random variable taking values on the interval (a,b). A function $f(x)$ is said to be the Probability density function of x if

i) $f(x) > 0 \ \forall \ x \in (a, b)$

ii) Total area under the probability curve is 1 $i.e, \int_a^b f(x)dx = 1.$

iii) For two distinct numbers 'c' and 'd' in (a, b) is given by

$P(c < x < d) = Area\ under\ the\ probability\ curve\ between\ ordinates\ x = c\ and\ x = d$

i. e $\int_c^d f(x)dx.$

Note: $P(c < x < d) = P(c \leq x \leq d) = P(c \leq x < d) = P(c < x \leq d)$

Cummulative distribution function of $f(x)$ is given by

$\int_{-\infty}^{x} f(x)dx$ i.e, $f(x) = \frac{d}{dx}F(x)$

Mean: The mean of the continuous Probability Distribution is defined as

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

Expectation: : The Expectation of the continuous Probability Distribution is defined as

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

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

Properties:

1) $E(X) = \mu$

2) $E(X) = k\, E(X)$

3) $E(X + k) = E(X) + k$

4)) $E(aX \pm b) = aE(X) \pm b$

Variance: The variance of the Continuous Probability Distribution is defined as

$$Var(X) = V(X) = \int_{-\infty}^{\infty} x^2\, f(x)dx - \mu^2.$$

Properties:

1) $V(c) = 0\ where\ c\ is\ a\ constant$

2) $V(kX) = k^2\, V(X)$

3) $V(X + k) = V(X)$

4) $V(aX \pm b) = a^2\, V(X)$

Mean Deviation: Mean deviation of continuous probability distribution function is defined as

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

Median: Median is the point which divides the entire distribution in to two equal parts. In case of continuous distribution,median is the point which divides the total area in to two equal parts i.e, $\int_a^M f(x)dx = \int_M^b f(x)dx = \frac{1}{2}\ \forall\ x \in (a, b)$.

Mode: Mode is the value of x for which $f(x)$ is maximum.

i.e $f'(x) = 0\ and\ f''(x) < 0$ for $x \in (a, b)$

PROBLEMS

1.If the probability density function $f(x) = \frac{k}{1+x^2}$ $-\infty < x < \infty$. Find the value of 'k' and probability distribution function of $f(x)$.

Sol: Since total area under the probability curve is 1 $i.e, \int_a^b f(x)dx = 1$.

$$\int_{-\infty}^{\infty} \frac{k}{1+x^2} dx = 1.$$

$$2k(\tan^{-1} x)_0^{\infty} = 1$$

$$2k(\tan^{-1} \infty - \tan^{-1} 0) = 1$$

$$\therefore k = \frac{1}{\pi}$$

C

Cummulative distribution function of $f(x)$ is given by

$$\int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} \frac{k}{1+x^2} dx = \frac{1}{\pi}(\tan^{-1} x)_{-\infty}^{x} = \frac{1}{\pi}[\frac{\pi}{2} + (\tan^{-1} x)].$$

2. If the probability density function $f(x) = ce^{-|x|}$ $-\infty < x < \infty$.

Find the value of 'c',mean and variance.

Sol: Since total area under the probability curve is 1 $i.e, \int_a^b f(x)dx = 1$.

$$\int_{-\infty}^{\infty} ce^{-|x|} dx = 1$$

$$2\int_0^{\infty} ce^{-x} dx = 1$$

$$2c\,(\frac{e^{-x}}{-1})_0^{\infty} = 1$$

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

$Mean = \mu = \int_{-\infty}^{\infty} x\, f(x)dx = \frac{1}{2}\int_{-\infty}^{\infty} xe^{-|x|} dx = 0$ *since* $xe^{-|x|}$ *is an odd function .*

$variance = V(X)$

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

$$= \frac{1}{2}\int_{-\infty}^{\infty} x^2 e^{-|x|} dx$$

$$= \frac{1}{2}\int_0^{\infty} 2x^2 e^{-x} dx = [x^2(-e^{-x}) - 2x(e^{-x}) + 2(-e^{-x})]_0^{\infty} = 2.$$

3. If the probability density function $f(x) = \begin{cases} \frac{sinx}{2} & if\ 0 \le x \le \pi \\ 0 & otherwise \end{cases}$.

Find mean,median,mode and $P(0 < x < \frac{\pi}{2})$.

Sol: *Mean*=$\mu = \int_{-\infty}^{\infty} x\, f(x)dx = \frac{1}{2}\int_0^{\pi} x\frac{sinx}{2}dx = \frac{1}{2}[-xcosx + sinx]_0^{\pi} = \frac{\pi}{2}$.

Let M be the Median then

$$\int_0^M f(x)dx = \int_M^{\pi} f(x)dx = \frac{1}{2}\ \forall\ x \in (-\infty, \infty)$$

$$\int_0^M \frac{\boldsymbol{sinx}}{\boldsymbol{2}}dx = \int_M^{\pi} \frac{\boldsymbol{sinx}}{\boldsymbol{2}}dx = \frac{1}{2}\ \forall\ x \in (-\infty, \infty)$$

consider $\int_M^{\pi} \frac{sinx}{2}dx = \frac{1}{2}$ then $(-cosx)_M^{\pi}$=1

$\therefore M = \frac{\pi}{2}$

Since $f(x) = \begin{cases} \frac{sinx}{2} & if\ 0 \le x \le \pi \\ 0 & otherwise \end{cases}$

To find maximum, we have $f'(x) = 0$

i.e, $cosx = 0\ implies\ that\ x = \frac{\pi}{2}$

and $f''(x) = -\frac{sinx}{2}\ which\ is\ less\ than\ 0\ at\ x = \frac{\pi}{2}$

$\therefore Mode = \frac{\pi}{2}$.

4.If the distributed function is given by

$$\boldsymbol{F(x) = \begin{cases} 0\ if\ x \le 1 \\ k(x-1)^4\ if\ 1 \le x \le 3 \\ 1\ if\ x > 3 \end{cases}}$$

Find $\boldsymbol{k, f(x), mean.}$

Sol: Cummulative distribution function of $f(x)$ is given by

$\int_{-\infty}^{x} f(x)dx$ i.e, $f(x) = \frac{d}{dx}F(x)$

i.e, $f(x) = \begin{cases} 0\ if\ x \le 1 \\ 4k(x-1)^3\ if\ 1 \le x \le 3 \\ 0\ if\ x > 3 \end{cases}$

Since total area under the probability curve is 1 $i.e, \int_a^b f(x)dx = 1$

$$\int_1^3 4k(x-1)^3 dx = 1$$

$$[k(x-1)^4]_1^3 = 1$$

$$\therefore k = \frac{1}{16}$$

$$\therefore f(x) = \begin{cases} 0 \text{ if } x \leq 1 \\ \frac{1}{4}(x-1)^3 \text{ if } 1 \leq x \leq 3 \\ 0 \text{ if } x > 3 \end{cases}$$

$$Mean = \mu = \int_{-\infty}^{\infty} x\, f(x) dx = \frac{1}{4}\int_1^3 x(x-1)^3 dx = 19.6.$$

Multiple Random variables

Discrete two dimensional random variable:

Joint probability mass function is defined as $f(x,y) = P(X = x_i, Y = y_i)$

Joint probability distribution function is defined as

$F_{XY}(x,y) = P(X < x_i, Y < y_i) = \sum_{<x}\sum_{<y} p(x_i, y_i)$

Marginal probability mass functions of X and Y are defined as

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

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

Continuous two dimensional random variable:

Joint probability density function is defined as

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

and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\, dxdy = 1$

Joint probability distribution function is defined as

$F_{XY}(x,y) = P(X < x_i, Y < y_i) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(x,y)\, dxdy$

and $f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y}[F_{XY}(x,y)]$

Marginal probability density functions of X is defined as

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy$$

Marginal probability density functions of Y is defined as

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

PROBLEMS

1.For the following 2-d probability distribution of X and Y

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

Find i) P(X$\leq 2, Y=2$) ii) $F_X(2)$ iii)P(Y=3) iv) P(X$< 3, Y \leq 4$) v) $F_y(3)$.

Sol: Given

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

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

= 0+0.12

= 0.12

ii) $F_X(2)$ =P(X≤ 2) = P(X$= 1$) + P(X$= 2$)

$=\sum_j p(x_i,y_j)+\sum_j p(x_i,y_j)$

=(0.1+0+0.2+0.1)+(0.05+0.2+0.08+0.1)

=0.66

iii) P(Y=3) = $\sum_i p(x_i,y_j)$

= 0.2+0.08+0.1

= 0.38.

iv) P(X$< 3, Y \leq 4$) = P(X$< 3, Y=1$)+ P(X$< 3, Y=2$)+ P(X$< 3, Y=3$)

+ P(X$< 3, Y=4$)

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

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

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

=(0.1+0+0.2+0.1)+(0.05+0.2+0.08+0.1)

=0.66

v) $F_y(3)$ = P(Y$\leq$ 3) = P(Y= 1) + P(Y= 2)+ P(Y= 3)

=(0.1+0.05+0.1)+(0+0.12+0.05)+(0.2+0.08+0.1)

=0.8

2.Suppose the random variables X and Y have the joint density function defined by

$$f(x,y) = \begin{cases} c(2x+y) \ if \ 2 < x < 6, 0 < y < 5 \\ 0 \quad otherwise \end{cases}$$

Find i)c ii)P(X>3,Y>2) iii) P(X>3)

Sol: Since $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1$

$$\int_2^6\int_0^5 c(2x+y)\,dydx = 1$$

$$\int_2^6 c(2xy+\frac{y^2}{2})_0^5\,dx = 1$$

$$\int_2^6 c(10x+\frac{25}{2})\,dx = 1$$

$$c(10\frac{x^2}{2}+\frac{25x}{2})_2^6 = 1$$

$$\therefore c = \frac{1}{210}$$

ii) $P(X > 3, Y > 2) = \int_3^6\int_2^5 f(x,y)\,dydx$

$$= \int_3^6\int_2^5 \frac{1}{210}(2x+y)\,dydx$$

$$= \frac{1}{210}\int_3^6 (2xy+\frac{y^2}{2})_2^5\,dx = \frac{15}{28}.$$

iii)) $P(X > 3) = \frac{1}{210}\int_3^6\int_0^5 f(x,y)\,dydx$

$$= \frac{1}{210}\int_3^6\int_0^5 (2x+y)\,dydx$$

$$= \frac{1}{210}\int_3^6 (2xy+\frac{y^2}{2})_0^5\,dx$$

$$= \frac{1}{210}\int_3^6 \left(10x + \frac{25}{2}\right) dx$$

$$= \frac{1}{210}\left[10x^2 + \left(10x + \frac{25x}{2}\right)\right]_3^6 = \frac{23}{28}.$$

3.The joint density function defined by

$$f(x,y) = \begin{cases} c(xy) \text{ if } 1 < x < 3, 2 < y < 4 \\ 0 \quad otherwise \end{cases}$$

Find i) c

ii) Marginal probability density functions of X and Y

iii) Show that X and Y are independent.

Sol: Since $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, dxdy = 1$

$$\int_2^4 \int_1^3 c(xy)dxdy = 1$$

$$\int_2^4 cy\left(\frac{x^2}{2}\right)_1^3 dy = 1$$

$$\frac{8c}{2}\left(\frac{y^2}{2}\right)_2^4 = 1 \quad \therefore c = \frac{1}{24}.$$

ii) Marginal probability density functions of X and Y

Marginal probability density functions of X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy = \frac{1}{24}\int_2^4 xy\, dy = \frac{x}{4}$$

Marginal probability density functions of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx = \frac{1}{24}\int_1^4 xy\, dx = \frac{y}{6}$$

iii)Clearly $f_{XY}(x,y) = \frac{xy}{24} == \frac{x}{4}\frac{y}{6} = f_X(x)\, f_Y(y)$

Therefore ,X and Y are independent.

Conditional probability density function :

Conditional probability density function of X on Y is

$$f_{XY}(X/Y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

Conditional probability density function of Y on X is

$$f_{XY}(Y/X) = \frac{f_{XY}(x,y)}{f_X(X)}$$

4.The joint density function defined by

$$f(x,y)=\begin{cases}(x^2+\frac{xy}{3}) \ if \ 0<x<1, 0<y<2\\ 0 \quad otherwise\end{cases}$$

Find

i) **Conditional probability density functions.**
ii) **Marginal probability density functions**
iii) **Check whether the functions X and Y are independent or not**

Sol: Given $f(x,y)=\begin{cases}(x^2+\frac{xy}{3}) \ if \ 0<x<1, 0<y<2\\ 0 \quad otherwise\end{cases}$

Marginal probability density functions of X is

$$f_X(x)=\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy=\int_0^2(x^2+\tfrac{xy}{3})\,dy=2x(x+\tfrac{1}{3})$$

Marginal probability density functions of Y is

$$f_Y(y)=\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx=\int_0^1(x^2+\tfrac{xy}{3})\,dx=\tfrac{1}{3}+\tfrac{y}{6}$$

$$Here \quad f_Y(y)\ f_X(x)=2x(x+\tfrac{1}{3})(\tfrac{1}{3}+\tfrac{y}{6})$$

Therefore, $f_{XY}(x,y)\neq f_X(x)\,f_Y(y)$

Hence X and Y are not Independent.

Conditional probability density function of X on Y is

$$f_{XY}(X/Y)=\frac{f_{XY}(x,y)}{f_Y(y)}=\frac{(x^2+\frac{xy}{3})}{(\frac{1}{3}+\frac{y}{6})}$$

Conditional probability density function of Y on X is

$$f_{XY}(Y/X)=\frac{f_{XY}(x,y)}{f_X(X)}=\frac{(x^2+\frac{xy}{3})}{2x(x+\frac{1}{3})}$$

UNIT- II

PROBABILITY DISTRIBUTIONS

BINOMIAL DISTRIBUTION: A Random variable 'X' has binomial distribution if it assumes only non-negative values with probability mass function given by

$$p(x=r)=\begin{cases} n_{c_r}P^r q^{n-r} & r=0,1,2,----n \\ 0 & otherwise \end{cases}$$

$= b(r;n,p)$

CONDITIONS FOR APPLICABILITY OF BINOMIAL DISTRIBUTIONS:

1. Number of trials must be finite (n is finite)
2. The trails are independent
3. There are only two possible outcomes in any event i.e., success and failure.
4. Probability of success in each trail remains constant.

Examples;

1. Tossing a coin n times
2. Throwing a die
3. No. of defective items in the box

MEAN OF THE BINOMIAL DISTRIBUTION

$$\mu = \sum_{r=0}^{n} r.P(r)$$

$$= \sum_{r=0}^{n} r.n_{c_r}P^r q^{n-r}$$

$$= n_{c_1}P^1 q^{n-1} + 2n_2 P^r q^{n-2} + 3n_{c_3}P^3 q^{n-3} + \cdots .... nn_{c_n}P^n q^{n-n}$$

$$= nP^1 q^{n-1} + 2.\frac{n(n-1)}{2!}p^2 q^{n-2} + 3.\frac{n(n-1)(n-2)}{3!}p^3 q^{n-3} + --+np^n$$

$$= np[\, q^{(n-1)} + (n-1)_{c_1}p^1 q^{(n-1)-1} + ----+ p^{n-1}$$

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

$$= np \qquad since\ [p+q=1]$$

Mean=np.

VARIANCE OF THE BINOMIAL DISTRIBUTION

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

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

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

$$= \sum_{r=0}^{n} r(r-1)n_{c_r}P^r q^{n-r} + np - n^2P^2$$

$$let \sum_{r=0}^{n} r(r-1)P(r) = \sum_{r=0}^{n} r(r-1) n_{c_r} P^r q^{n-r} = 2\, n_{c_{2r}} P^2 q^2 n^{n-2} + 6 n_{c_3} P^3 q^{n-3} + 12 n_{c_r} P^4 q^{n-4} + \cdots + n(n-1)\, P^n$$

$$= n(n-1)P^2 \left[q^{n-2} + +(n-2)_{c_1} p^1 q^{(n-2)-1} + \cdots + p^2 \right]$$

$$= n(n-1)P^2 (p+q)^{n-2}$$

$$= n^2 P^2 - nP^2$$

$$\sigma^2 = n^2 P^2 - nP^2 + np - n^2 P^2$$

$$= np(1-p)$$

$$= npq.$$

PROBLEMS

1.In tossing a coin 10 times simultaneously.Find the probability of getting

i)atleast 7 heads ii) atmost 3 heads iii)exactly 6 heads.

SOL: Given $n = 10$

Probability of getting a head in tossing a coin = $\frac{1}{2} = p$.

Probability of getting no head = $q = 1 - \frac{1}{2} = \frac{1}{2}$.

The probability of getting r heads in a throw of 10 coins is

$$P(X = r) = p(r) = 10_{C_r} \left(\frac{1}{2}\right)^r \left(\frac{1}{2}\right)^{10-r} ; r = 0,1,2, \ldots \ldots, 10$$

(i) Probability of getting atleast seven heads is given by

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

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

$$= \frac{1}{2^{10}} \left[10_{C_7} + 10_{C_8} + 10_{C_9} + 10_{C_{10}} \right]$$

$$= \frac{1}{2^{10}} [120 + 45 + 10 + 1]$$

$$= \frac{176}{1024}$$

$$= 0.1719$$

ii) Probability of getting atmost 3 heads is given by

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

$$= 10_{C_1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{10-1} + 10_{C_2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{10-2} + 10_{C_3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{10-3} + 10_{C_0}\left(\frac{1}{2}\right)^{10}$$

$$= \frac{1}{2^{10}}\left[10_{C_0} + 10_{C_1} + 10_{C_2} + 10_{C_3}\right]$$

$$= \frac{1}{2^{10}}[120 + 45 + 10 + 1]$$

$$= \frac{176}{1024}$$

$$= 0.1719$$

iii)Probability of getting exactly six heads is given by

$$P(X=6) = 10_{C_6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^{10-6}$$

=0.205.

2.In 256 sets of 12 tosses of a coin ,in how many cases one can expect 8 Heads and 4 Tails.

Solution: The probability of getting a head, $p = \frac{1}{2}$

The probability of getting a tail, $q = \frac{1}{2}$

Here $n = 12$

The probability of getting 8heads and 4Tails in 12trials $= P(X=8) = 12_{C_8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^4$

$$= \frac{12!}{8!\,4!}\left(\frac{1}{2}\right)^{12} = \frac{495}{2^{12}}$$

The expected number of getting 8 heads and 4 Tails in 12 trials of such cases in256 sets

$$= 256 \times P(X=8) = 2^8 \times \frac{495}{2^{12}} = \frac{495}{16} = 30.9375 \sim 31$$

3.Find the probability of getting an even number 3 or 4 or 5 times in throwing a die 10 times.

Sol: Probability of getting an even number in throwing a die $= \frac{3}{6} = \frac{1}{2} = p.$

Probability of getting an odd number in throwing a die $= q = \frac{1}{2}.$

∴Probability of getting an even number 3 or 4 or 5 times in throwing a die 10 times is

$$P(X=3) + P(X=4) + P(X=5)$$

$$= 10_{C_3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{10-3} + 10_{C_4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{10-4} + 10_{C_5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^{10-5}$$

$$= \frac{1}{2^{10}}\left[10_{C_3} + 10_{C_4} + 10_{C_5}\right]$$

$$= \frac{1}{2^{10}}\left[120 + 252 + 210\right]$$

=0.568.

4.Out of 800 families with 4 children each,how many could you expect to have

a)three boys b)five girls c) 2 or 3 boys d)atleast 1 boy.

Sol: : Given $n = 5, N = 800$

Let having boys be success

Probability of having a boy $= \frac{1}{2} = p.$

Probability of having girl $= q = 1 - \frac{1}{2} = \frac{1}{2}.$

The probability of having r boyss in 5 children is

$$P(X = r) = p(r) = 5_{C_r}\left(\frac{1}{2}\right)^r\left(\frac{1}{2}\right)^{5-r}; r = 0,1,2 \ldots\ldots 5$$

a)Probability of having 3 boys is given by

$$P(X = 3) = 5_{C_r}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{5-3} = \frac{5}{16}$$

Expected number of families having 3 boys = $N\,p(3)$ =800($\frac{5}{16}$) =250 families.

b) Probability of having 5 girls = Probability of having no boys is given by

$$P(X = 0) = 5_{C_0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^{5-0} = \frac{1}{32}$$

Expected number of families having 5 girls = $N\,p(0)$ =800($\frac{1}{32}$) =25 families.

c) Probability of having either 2 or 3 boys is given by

$$P(X = 2) + P(X = 3) = 5_{C_2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{5-2} + 5_{C_3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{5-3} = \frac{5}{18}$$

Expected number of families having 3 boys = $N\,p(3)$ =800($\frac{5}{8}$) =500 families.

d) Probability of having atleast 1 boy is given by

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

$$= 1 - 5_{C_0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^{5-0} = \frac{31}{32}$$

Expected number of families having atleast 1 boy =800($\frac{31}{32}$) =775 families.

5.Fit a Binomial distribution for the following data.

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

Sol: Given n= 5,$\sum f = 2 + 14 + 20 + 34 + 22 + 8 = 100$

$\sum x_i f_i$== $0(2) + 1(14) + 2(20) + 3(34) + 4(22) + 5(8)$ = 284

$\therefore$ *Mean of the distribution* $= \frac{\sum x_i f_i}{\sum f_i}$

$= \frac{284}{100}$ = 2.84.

We have Mean of the binomial distribution = np = 2.84

$\therefore p = \frac{2.84}{5} = 0.568; q = 1 - 0.568 = 0.432.$

TABLE TO FIT BINOMIAL DISTRIBUTION

X	P(x_i)	E(x_i)
0	$5_{C_0}(0.568)^0\,(0.432)^{5-0}$=0.02	N p(0) =100(0.02)=2
1	$5_{C_1}(0.568)^1\,(0.432)^{5-1}$=0.09	9
2	$5_{C_2}(0.568)^2\,(0.432)^{5-2}$=0.26	26
3	$5_{C_3}(0.568)^3\,(0.432)^{5-3}$=0.34	34

4	$5_{C_4}(0.568)^4\ (0.432)^{5-4}$=0.22	22
5	$5_{C_5}(0.568)^5\ (0.432)^{5-5}$=0.059	5.9

Fitted Binomial distribution is

x	0	1	2	3	4	5
f	2	10	26	34	22	6

RECURRENCE RELATION

$p(r+1) = n_{C_{r+1}}(p)^{r+1}\ (q)^{n-r-1}$ ……(1)

$p(r) = n_{C_r}(p)^r\ (q)^{n-r}$ ……(2)

$\frac{(1)}{(2)} = \frac{p(r+1)}{p(r)} = \frac{n_{C_{r+1}}(p)^{r+1}(q)^{n-r-1}}{n_{C_r}(p)^r(q)^{n-r}}$

$\therefore \frac{p(r+1)}{p(r)} = \frac{n_{C_{r+1}}}{n_{C_r}}\left(\frac{p}{q}\right)$

$p(r+1) = \frac{n_{C_{r+1}}}{n_{C_r}}\left(\frac{p}{q}\right)p(r).$

POISSON DISTRIBUTION:

A random variable 'X' follows Poisson distribution if it assumes only non-negative values with probability mass function is given by

$$P(x=r) = P(r,\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^r}{r!} & for \quad y = 0,1,--(\lambda > 0) \\ 0 & otherwise \end{cases}$$

CONDITIONS FOR POISSON DISTRIBUTION:

1. The number of trials are very large (infinite)
2. The probability of occurrence of an event is very small $(\lambda = np)$
3. $\lambda = np = finite$

Examples:

1. The number of printing mistakes per page in a large text

2. The number of telephone calls per minute at a switch board
3. The number of defective items manufactured by a company.

RECURRENCE RELATION:

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

$P(r) = \frac{e^{-\lambda}\lambda^{r}}{(r)!}$ -------(2)

$$\frac{1}{2} = \frac{P(r+1)}{P(r)} = \frac{e^{-\lambda}\lambda^2.\lambda}{(r+1)r!} X \frac{r!}{e^{-\lambda}\lambda^2}$$

$$P(r+1) = \left(\frac{\lambda}{r+1}\right) P(r) \; for \; r = 0,1,2 - - - -$$

PROBLEMS

1.Using Recurrence relation find probability when x=0,1,2,3,4,5, if mean of P.D is 3.

Sol: We have

$P(r+1) = \left(\frac{\lambda}{r+1}\right) P(r) \; for \; r = 0,1,2 - - - -$(1)

Given λ= 3

$P(0) = \frac{e^{-3}\lambda^{0}}{(0)!} = e^{-3}$ [by definition of poisson distribution]

From (1),

For $r = 0$, $P(1) = \left(\frac{3}{0+1}\right) P(0)$ =3 e^{-3}

For $r = 1$, $P(2) = \left(\frac{3}{1+1}\right) P(0) = \frac{3}{2} e^{-3}$

For $r = 2$, $P(3) = \left(\frac{3}{2+1}\right) P(0) = e^{-3}$

F For $r = 3$, $P(4) = \left(\frac{3}{3+1}\right) P(0) = \frac{3}{4} e^{-3}$

or $r = 4$, $P(5) = \left(\frac{3}{4+1}\right) P(0) = \frac{3}{5} e^{-3}$.

2.If X is a random variable such that $3P(X = 4) = \frac{P(X=2)}{2} + P(X = 0)$.

Find mean, $P(X \leq 2)$.

Sol: Given $3P(X = 4) = \frac{P(X=2)}{2} + P(X = 0)$…..(1)

Since X is a poisson variable,

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

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

Solving it we get $\lambda^4 - 2\lambda^2 - 4 = 0$

Taking $\lambda^2 = k, we\ get\ k^2 - 2k - 4 = 0$

$\therefore k = 4, -2$

$\therefore \lambda^2 = 4\ implies\ that\ \lambda = 2$

Therefore, Mean of the poisson distribution = 2

$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$$= \frac{e^{-2}\lambda^0}{0!} + \frac{e^{-2}2^1}{1!} + \frac{e^{-\lambda}2^2}{2!} = 0.54.$$

3.A car hire firm has 2 cars which it hires out day by day.The number of demands for a car on each day is distributed as poisson with mean 1.5 Calculate the proportion of days

i)on which there is no demand

ii) on which demand is refused.

Sol:Let number of demands for cars be the success.

Given mean = 1.5= λ

Using poisson distribution,

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

i)Probability that there is no demand for car is

$$P(x = 0) = \frac{e^{-1.5}(1.5)^0}{0!} = 0.223$$

Expected number of days that there is no demand =N

$P(0) = 365(0.223) = 81.39 \sim 81\ days$

ii) Probability that demand refused for car is

$P(x > 2) = 1 - P(x = 0) - P(x = 1) - P(x = 2)$

$$= 1 - \frac{e^{-1.5}(1.5)^0}{0!} - \frac{e^{-1.5}(1.5)^1}{1!} - \frac{e^{-1.5}(1.5)^2}{2!} = 0.191$$

Expected number of days that demand refused for car = N$P(x > 2)$

$= 365(0.191) = 69.7 \sim 70\ days.$

4.The distribution of typing mistakes committed by typist is given below.

Fit a Poisson distribution for it.

Mistakes per page	0	1	2	3	4	5
Number of pages	142	156	69	27	5	1

Sol: Given n= 5, $\sum f = 142 + 156 + 69 + 27 + 5 + 1 = 400$

$\sum x_i f_i == 0(142) + 1(156) + 2(69) + 3(27) + 4(5) + 5(1) = 400$

$\therefore Mean\ of\ the\ distribution = \frac{\sum x_i f_i}{\sum f_i}$

$= \frac{400}{400} = 1.$

We have Mean of the poisson distribution = $\lambda = 1$

TABLE TO FIT POISSON DISTRIBUTION

X	P(x_i)	E(x_i)
0	$\frac{e^{-1}(1)^0}{0!}$=0.368	N p(0) =400(0.368)=147.2~147
1	$\frac{e^{-1}(1)^1}{1!}$=0.368	147
2	$\frac{e^{-1}(1)^2}{2!}$=0.184	74
3	$\frac{e^{-1}(1)^3}{3!}$=0.061	24
4	$\frac{e^{-1}(1)^4}{4!} = 0.015$	6
5	$\frac{e^{-1}(1)^5}{5!} = 0.003$	1

Fitted Poisson distribution is

Mistakes per page	0	1	2	3	4	5
Number of pages	147	147	74	24	6	1

NORMAL DISTRIBUTION (GAUSSIAN DISTRIBUTION)

Let X be a continuous random variable, then it is said to follow normal distribution if its pdf is given by

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty \le x \le \infty, \mu, \sigma > 0$$

Here , σ are the mean & S.D of X.

PROPERTIES OF NORMAL DISTRIBUTION:

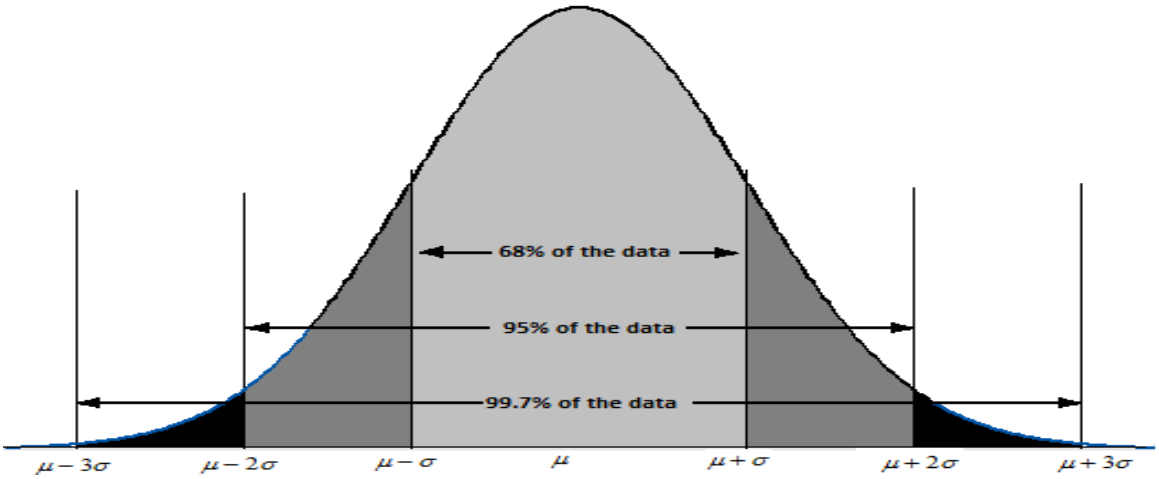
1. Normal curve is always centered at mean
2. Mean, median and mode coincide (i.e., equal)
3. It is unimodal
4. It is a symmetrical curve and bell shaped curve
5. X-axis is an asymptote to the normal curve
6. The total area under the normal curve from $-\infty \; to \infty$ is "1"
7. The points of inflection of the normal curve are $\mu \pm \sigma, \mu \pm 3\sigma$
8. The area of the normal curve between

$$\mu - \sigma \text{ to } \mu + \sigma = 68.27\%$$

$$\mu - 2\sigma \; to \; \mu + 2\sigma = 95.44\%$$

$$\mu - 3\sigma \; to \; \mu + 3\sigma = 99.73\%$$

9.The curve for normal distribution is given below



STANDARD NORMAL VARIABLE

Let $Z = \frac{x-\mu}{\sigma}$ with mean '0' and variance is '1' then the normal variable is said to be standard normal variable.

STANDARD NORMAL DISTRIBUTION

The normal distribution with man '0' and variance '1' is said to be standard normal distribution of its probability density function is defined by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x \leq \infty$$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad -\infty \leq x \leq \infty \ \ (\mu = 0, \sigma = 1)$$

MEAN OF THE NORMAL DISTRIBUTION

Consider Normal distribution with b,σ as parameters Then

$$f(x; b, \sigma) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-b)^2}{2\sigma^2}}$$

Mean of the continuous distribution is given by

$$\mu = \int_{-\infty}^{\infty} x\, f(x)dx = \int_{-\infty}^{\infty} x\, \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-b)^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + b)\, e^{-\frac{(z)^2}{2}} dz$$

$[Putting\ z = \frac{x-b}{\sigma}\ \ so\ that\ dx = \sigma\, dz]$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z\, e^{-\frac{(z)^2}{2}} dz + \frac{b}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(z)^2}{2}} dz$$

$$=\frac{2b}{\sqrt{2\pi}}\int_{-0}^{\infty} e^{-\frac{(z)^2}{2}}dz$$

[since $z\,e^{-\frac{(z)^2}{2}}$ *is an odd function and* $e^{-\frac{(z)^2}{2}}$ *is an even function*]

$$\mu = \frac{2b}{\sqrt{2\pi}}\sqrt{\frac{\pi}{2}} \quad = b$$

$\therefore Mean = \ b$

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

VARIANCE OF THE NORMAL DISTRIBUTION

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

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx - \mu^2$$

Let $z = \frac{x-\mu}{\sigma} \Longrightarrow dx = \sigma dz$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} (\mu^2 + \sigma^2 z^2 \ + 2\mu\sigma z)\, e^{-\frac{z^2}{2}}\sigma dz - \mu^2$$

$$= \frac{\mu^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}dz + \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2 e^{-\frac{s^2}{2}}dz + \frac{2\mu\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2 e^{-\frac{s^2}{2}}dz - \mu^2$$

$$= \frac{2\mu^2}{\sqrt{2\pi}}\int_{0}^{\infty} e^{-\frac{z^2}{2}}dz + \frac{2\sigma^2}{\sqrt{2\pi}}\int_{0}^{\infty} z^2 e^{-\frac{s^2}{2}}dz - \mu^2$$

$$= \frac{2\sigma^2}{\sqrt{2\pi}}\int_0^{\infty} z^2 e^{-\frac{z^2}{2}}dz$$

$\because \frac{z^2}{2} = + \Rightarrow \frac{2zdz}{2} = dt$ $\qquad dz = \frac{dt}{\sqrt{2t}}$

$$= \frac{2\sigma^2}{\sqrt{2\pi}}\int_{0}^{\infty} (2+)^2 e^t \frac{dt}{\sqrt{2t}}$$

$$= \frac{2\sigma^2}{\sqrt{\pi}}\int_0^{\infty} e^{-t} + \frac{s}{2} - 1.\text{dt}$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right)$$

$$== \frac{2\sigma^2}{\sqrt{\pi}} \frac{1}{2} \Gamma\left(\frac{1}{2}\right)$$

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

MEDIAN OF THE NORMAL DISTRIBUTION

Let x=M be the median, then

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

Let $\mu \epsilon (-\infty, M)$

Let $\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\mu} f(x)dx + \int_{\mu}^{M} f(x)dx = \frac{1}{2}$

Consider $\int_{-\infty}^{\mu} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$

Let $z = \frac{x-\mu}{\sigma} \Rightarrow dx = \sigma dz$ $\quad [\because Limits\ of\ z\ -\infty \longrightarrow 0]$

$\int_{-\infty}^{\mu} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{z^2}{2}} \sigma dz$

$= \frac{1}{\sqrt{2\pi}} \int_{\infty}^{0} e^{-\frac{t^2}{2}} (dt)$ (by taking z=-t again)

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

From (1)

$$\int_{\mu}^{\mu} f(x)dx = 0 \Rightarrow \mu = M$$

MODE OF THE NORMAL DISTRIBUTION

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} - \left(\frac{x-\mu}{\sigma}\right)^2$$

$$f`(x) = 0 \Rightarrow \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \left(\frac{-1}{2}\right) 2\left(\frac{x-\mu}{\sigma}\right) \frac{1}{\sigma} = 0$$

$\Rightarrow x-\mu=0 \Rightarrow x=\mu$

$\Rightarrow x=\mu$

$$f^{11}(x)=\frac{-1}{\sigma^3\sqrt{2\pi}}\left[e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.1+(x-\mu)e^{-\frac{1}{2}}\left(\frac{x-\mu}{\sigma}\right)^2\left(\frac{-1}{2}\right)2\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}\right]$$

$$=\frac{-1}{\sigma^3\sqrt{2\pi}}\ [e^0+0]$$

$=\frac{-1}{\sigma^3\sqrt{2\pi}}<0$

$\because x=\mu$ *is the mode of normal distribution.*

1.If X is a normal variate, find the area A

i) to the left of $z=1.78$

ii) to the right of $z=-1.45$

iii) Corresponding to $-0.8\le z\le 1.53$

iv) to the left of $z=-2.52$ and to the right of $z=1.83$.

Sol: i) $P(z<-1.78)=0.5-P(-1.78<z<0)$

$=0.5-P(0<z<1.78)$

= 0.5-0.4625 =0.0375.

ii) $P(z>-1.45)=0.5+P(-1.45<z<0)$

$=0.5+P(0<z<1.45)$

= 0.5+0.4625 =0.9265.

iii) $P(-0.8\le z\le 1.53)=P(-0.8\le z\le 0)+P(0\le z\le 1.53)$

= 0.2881+0.4370=0.7251.

iv) $P(z<-2.52)=0.5-P(0<z<2.52)$=0.0059

$P(z>1.83)$ $=0.5-P(0<z<1.83)$

=0.036

2.If the masses of 300 students are normally distributed with mean 68 kgs and standard deviation 3kgs.How many students have masses

i)greater than 72kgs.

ii)less than or equal to 64 kgs

iii)between 65 and 71 kgs inclusive.

Sol:Given N=300,$\mu = 68, \sigma = 3$.Let X be the masses of the students.

i) Standard normal variate for X=72 is

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

$P(X > 72) = P(z > 1.33)$

$== 0.5 - P(0 < z < 1.33)$

=0.5-0.4082

=0.092

Expected number of students greater than 72 = E(X>72)

=300(0.092)

=27.54~28 students

II) Standard normal variate for X=64 is

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

$P(X \leq 64) = P(z \leq -1.33)$

$== 0.5 - P(0 < z < 1.33)$ (Using symmetry)

=0.5-0.4082

=0.092

Expected number of students less than or equal to 64 = E(X less than or equal to 64)

=300(0.092)

=27.54~28 students .

iii)Standard normal variate for X=65 is

$$z_1 = \frac{x-\mu}{\sigma} = \frac{65-68}{3} = -1$$

Standard normal variate for X=71 is

$$z_2 = \frac{x-\mu}{\sigma} = \frac{71-68}{3} = 1$$

$P(65 \leq X \leq 71) = P(-1 \leq z \leq 1)$

$= P(-1 \leq z \leq 0) + P(-0 \leq z \leq 1)$

$=2\ P(-0 \le z \le 1)$

$=2(0.341)= 0.6826$

$E(65 \le X \le 71) = 300(0.6826) = 205\ Students.$

∴Expected number of students between 65 and 71 kgs inclusive = 205 students.

3.In a normal distribution 31% of the items are under 45 and 8% of the items are over 64. Find mean and variance of the distribution.

Sol: Given $P(X < 45) = 31\% = 0.31$

And $P(X > 64) = 8\% = 0.08$

Let Mean and variances of the normal distributions are μ, σ^2.

Standard normal variate for X is

$$z = \frac{x-\mu}{\sigma}$$

Standard normal variate for X_1=45 is

$$z_1 = \frac{X_1-\mu}{\sigma} = \frac{45-\mu}{\sigma}$$

$$\Rightarrow \mu + \sigma z_1 = 45 \dots\dots\dots (1)$$

Standard normal variate for X_2=64 is

$$z_2 = \frac{X_2-\mu}{\sigma} = \frac{64-\mu}{\sigma}$$

$$\Rightarrow \mu + \sigma z_2 = 64 \dots\dots\dots (2)$$

From normal curve ,we $have\ P(-z_1 \le z \le 0) = 0.19$

$\Rightarrow z_1 = -0.5$

$P(0 \le z \le z_2)$=0.42

$\Rightarrow z_2 = 1.41$

$substituting\ the\ values\ of\ z_1, z_2$ in (1) and (2),we get

$\mu = 50, \sigma^2 = 98.$

4. In a normal distribution 7% of the items are under 35 and 89% of the items are under 63. Find mean and variance of the distribution.

Sol: Given $P(X < 35) = 7\% = 0.07$

And $P(X < 63) = 89\% = 0.89$

Let Mean and variances of the normal distributions are μ, σ^2.

Standard normal variate for X is

$$z = \frac{x-\mu}{\sigma}$$

Standard normal variate for X_1=35 is

$$z_1 = \frac{X_1-\mu}{\sigma} = \frac{35-\mu}{\sigma}$$
$$\Rightarrow \mu + \sigma z_1 = 35 \dots\dots\dots (1)$$

Standard normal variate for X_2=63 is

$$z_2 = \frac{X_2-\mu}{\sigma} = \frac{63-\mu}{\sigma}$$
$$\Rightarrow \mu + \sigma z_2 = 63 \dots\dots\dots (2)$$

Given $P(X < 35) = P(z < z_1)$

0.07 = 0.5- $P(-z_1 \le z \le 0)$

$P(0 \le z \le z_1) = 0.43$

From normal curve ,we *have*

$\Rightarrow z_1 = 1.48$

We have $P(X < 63) = P(z < z_2)$

0.89 = 0.5+$P(0 \le z \le z_2)$

$P(0 \le z \le z_2) = 0.39$

From normal curve ,we *have*

$\Rightarrow z_2 = 1.23$

substituting the values of z_1, z_2 in (1) and (2),we get

$\mu = 50, \sigma^2 = 100$.

UNIT –III

SAMPLING

Introduction: The totality of observations with which we are concerned , whether this number be finite or infinite constitute population. In this chapter we focus on sampling from distributions or populations and such important quantities as the sample mean and sample variance.

Def: Population is defined as the aggregate or totality of statistical data forming a subject of investigation .

EX. The population of the heights of Indian.

The number of observations in the population is defined to be the size of the population. It may be finite or infinite .Size of the population is denoted by N.As the study of entire population may not be possible to carry out and hence a part of the population alone is selected.

Def: A portion of the population which is examined with a view to determining the population characteristics is called a sample . In other words, sample is a subset of population. Size of the sample is denoted by n.

The process of selection of a sample is called Sampling. There are different methods of sampling

- Probability Sampling Methods
- Non-Probability Sampling Methods

Probability Sampling Methods:

a) Random Sampling (Probability Sampling):

It is the process of drawing a sample from a population in such a way that each member of the population has an equal chance of being included in the sample.

Ex: A hand of cards from a well shuffled pack of cards is a random sample.

Note : If N is the size of the population and n is the size of the sample, then

- The no. of samples with replacement = N^n
- The no. of samples without replacement = N_{C_n}

b) Stratified Sampling :

In this , the population is first divided into several smaller groups called strata according to some relevant characteristics . From each strata samples are selected at random, all the samples are combined together to form the stratified sampling.

c) Systematic Sampling (Quasi Random Sampling):

In this method , all the units of the population are arranged in some order . If the population size is N, and the sample size is n, then we first define sample interval denoted by $= \frac{N}{n}$. then from first k items ,one unit is selected at random. Then from first unit every k^{th} unit is serially selected combining all the selected units constitute a systematic sampling.

Non Probability Sampling Methods:

a) Purposive (Judgment) Sampling :

In this method, the members constituting the sample are chosen not according to some definite scientific procedure , but according to convenience and personal choice of the individual who selects the sample . It is the choice of the individual items of a sample entirely depends on the individual judgment of the investigator.

b) **Sequential Sampling:**

It consists of a sequence of sample drawn one after another from the population. Depending on the results of previous samples if the result of the first sample is not acceptable then second sample is drawn and the process continues to take proper decision . But if the first sample is acceptable ,then no new sample is drawn.

Classification of Samples:

- **Large Samples :** If the size of the sample n ≥ 30 , then it is said to be large sample.
- **Small Samples :** If the size of the sample n < 30 ,then it is said to be small sample or exact sample.

Parameters and Statistics:

Parameter is a statistical measure based on all the units of a population. Statistic is a statistical measure based on only the units selected in a sample.

Note :In this unit , Parameter refers to the population and Statistic refers to sample.

Central Limit Theorem: If $\bar{x}$ be the mean of a random sample of size n drawn from population having mean μ and standard deviation σ , then the sampling distribution of the sample mean $\bar{x}$ is approximately a normal distribution with mean μ and SD = S.E of $\bar{x} = \frac{\sigma}{\sqrt{n}}$ provided the sample size n is large.

Standard Error of a Statistic : The standard error of statistic 't' is the standard deviation of the sampling distribution of the statistic i.e, S.E of sample mean is the standard deviation of the sampling distribution of sample mean.

Formulae for S.E:

- S.E of Sample mean $\bar{x} = \frac{\sigma}{\sqrt{n}}$ i.e, S.E $(\bar{x}) = \frac{\sigma}{\sqrt{n}}$
- S.E of sample proportion p= $\sqrt{\frac{PQ}{n}}$ i.e, S.E (p) = $\sqrt{\frac{PQ}{n}}$ where Q=1-P
- S.E of the difference of two sample means $\overline{x_1}$ and $\overline{x_2}$ i.e, S.E $(\overline{x_1} - \overline{x_2}) = \sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}$
- S.E of the difference of two proportions i.e, S.E$(p_1 - p_2) = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$

Estimation :

To use the statistic obtained by the samples as an estimate to predict the unknown parameter of the population from which the sample is drawn.

Estimate : An estimate is a statement made to find an unknown population parameter.

Estimator : The procedure or rule to determine an unknown population parameter is called estimator.

Ex. Sample proportion is an estimate of population proportion , because with the help of sample proportion value we can estimate the population proportion value.

Types of Estimation:

- **Point Estimation:** If the estimate of the population parameter is given by a single value , then the estimate is called a point estimation of the parameter.
- **Interval Estimation:** If the estimate of the population parameter is given by two different values between which the parameter may be considered to lie, then the estimate is called an interval estimation of the parameter.

Confidence interval Estimation of parameters:

In an interval estimation of the population parameter θ, if we can find two quantities t_1 and t_2 based on sample observations drawn from the population such that the unknown parameter θ is included in the interval $[t_1, t_2]$ in a specified cases ,then this is called a confidence interval for the parameter θ.

Confidence Limits for Population mean μ

- 95% confidence limits are $\bar{x} \pm 1.96\ (S.E.of\ \bar{x})$
- 99% confidence limits are $\bar{x} \pm 2.58\ (S.E.of\ \bar{x})$
- 99.73% confidence limits are $\bar{x} \pm 3\ (S.E.of\ \bar{x})$

- 90% confidence limits are $\bar{x} \pm 1.645\ (S.E.\ of\ \bar{x})$

Confidence limits for population proportion P

- 95% confidence limits are $p \pm 1.96$(S.E.of p)
- 99% confidence limits are $p \pm 2.58$(S.E. of p)
- 99.73% confidence limits are $p \pm 3$(S.E.of p)
- 90% confidence limits are $p \pm 1.645$(S.E.of p)

Confidence limits for the difference of two population means μ_1 and μ_2

- 95% confidence limits are $((\overline{x_1} - \overline{x_2}) \pm 1.96$ (S.E of $((\overline{x_1} - \overline{x_2}))$
- 99% confidence limits are $((\overline{x_1} - \overline{x_2}) \pm 2.58$ (S.E of $((\overline{x_1} - \overline{x_2}))$
- 99.73% confidence limits are $((\overline{x_1} - \overline{x_2}) \pm 3$ (S.E of $((\overline{x_1} - \overline{x_2}))$
- 90% confidence limits are $((\overline{x_1} - \overline{x_2}) \pm 2.58$ (S.E of $((\overline{x_1} - \overline{x_2}))$

Confidence limits for the difference of two population proportions

- 95% confidence limits are p_1-$p_2 \pm 1.96$ (S.E. of p_1-p_2)
- 99% confidence limits are p_1-$p_2 \pm 2.58$ (S.E. of p_1-p_2)
- 99.73% confidence limits are p_1-$p_2 \pm 3$ (S.E. of p_1-p_2)
- 90% confidence limits are p_1-$p_2 \pm 1.645$ (S.E. of p_1-p_2)

Determination of proper sample size

Sample size for estimating population mean :

n= $\left(\frac{z\sigma}{E}\right)^2$ where z – Level of significance

σ – Standard deviation of population and

E – Maximum sampling Error = $\bar{x} - \mu$

Sample size for estimating population proportion :

$n = \frac{z^2 PQ}{E^2}$ where z – Level of significance

P – Population proportion

Q – 1-P

E – Maximum Sampling error = p-P

Testing of Hypothesis :

It is an assumption or supposition and the decision making procedure about the assumption whether to accept or reject is called hypothesis testing .

Def: Statistical Hypothesis : To arrive at decision about the population on the basis of sample information we make assumptions about the population parameters involved such assumption is called a statistical hypothesis .

Procedure for testing a hypothesis:

Test of Hypothesis involves the following steps:

Step1: Statement of hypothesis :

There are two types of hypothesis :

- **Null hypothesis:** A definite statement about the population parameter. Usually a null hypothesis is written as no difference , denoted by H_0.

  Ex. $H_0: \mu = \mu_0$
- **Alternative hypothesis :** A statement which contradicts the null hypothesis is called alternative hypothesis. Usually an alternative hypothesis is written as some difference , denoted by H_1.

  Setting of alternative hypothesis is very important to decide whether it is two-tailed or one – tailed alternative , which depends upon the question it is dealing.

  Ex.$H_1: \mu \neq \mu_0$ (Two – Tailed test)

  or

  $H_1: \mu > \mu_0$ (Right one tailed test)

  or

  $H_1: \mu < \mu_0$ (Left one tailed test)

Step 2: Specification of level of significance :

The LOS denoted by α is the confidence with which we reject or accept the null hypothesis. It is generally specified before a test procedure ,which can be either 5% (0.05) , 1% or 10% which means that thee are about 5 chances in 100 that we would reject the null hypothesis H_0 and the remaining 95% confident that we would accept the null hypothesis H_0 . Similarly , it is applicable for different level of significance.

Step 3 : Identification of the test Statistic :

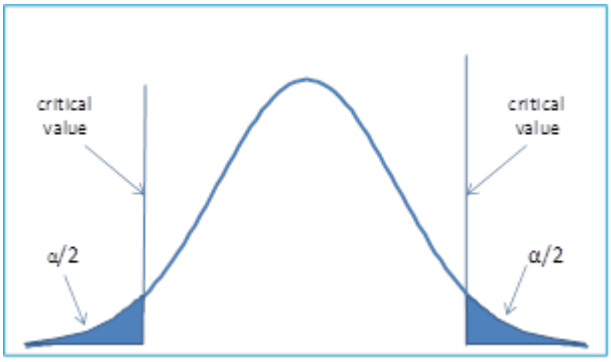
There are several tests of significance like z,t, F etc .Depending upon the nature of the information given in the problem we have to select the right test and construct the test criterion and appropriate probability distribution.

Step 4: Critical Region:

It is the distribution of the statistic .

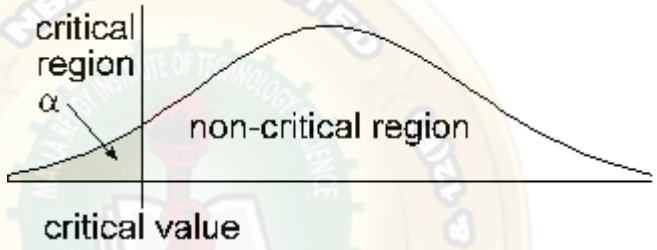
- **Two – Tailed Test : The critical region under the curve is equally distributed on both sides of the mean.**

  If H_1 has $\neq$ sign , the critical region is divided equally on both sides of the distribution.

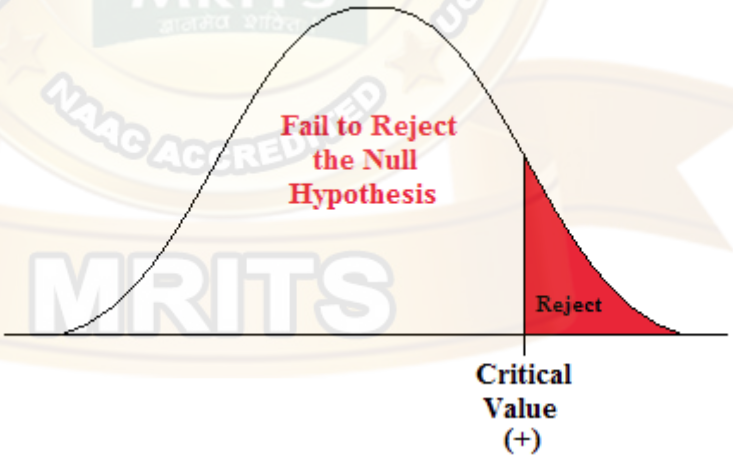


- **One Tailed Test: The critical region under the curve is distributed on one side of the mean.**

Left one tailed test: If H_1 has $<$ sign , the critical region is taken in the left side of the distribution.



Right one tailed test : If H_1 has $>$ sign , the critical region is taken on right side of the distribution.



Step 5 : Making decision:

By comparing the computed value and the critical value decision is taken for accepting or rejecting H_0

If calculated value $\leq$ critical value , we accept H_0, otherwise reject H_0.

Errors of Sampling :

While drawing conclusions for population parameters on the basis of the sample results , we have two types of errors.

- **Type I error :** Reject H_0 when it is true i.e, if the null hypothesis H_0 is true but it is rejected by test procedure .
- **Type II error :** Accept H_0 when it is false i.e, if the null hypothesis $H_0$is false but it is accepted by test procedure.

DECISION TABLE

	H_0 is accepted	**H_0 is rejected**
H_0 is true	**Correct Decision**	**Type I Error**
H_0 is false	**Type II Error**	**Correct Decision**

Problems:

1. If the population is 3,6,9,15,27
   a) List all possible samples of size 3 that can be taken without replacement from finite population
   b) Calculate the mean of each of the sampling distribution of means
   c) Find the standard deviation of sampling distribution of means

   Sol: Mean of the population , $\mu = \frac{3+6+9+15+27}{5} = \frac{60}{5} = 12$

   Standard deviation of the population ,

$$\sigma = \sqrt{\frac{(3-12)^2+(6-12)^2+(9-12)^2+(15-12)^2+(27-12)^2}{5}}$$

$$= \sqrt{\frac{81+36+9+9+225}{5}} = \sqrt{\frac{360}{5}} = 8.4853$$

   a) Sampling without replacement :
   The total number of samples without replacement is $N_{C_n} = 5_{C_s} = 10$

   The 10 samples are (3,6,9), (3,6,15), (3,9,15), (3,6,27), (3,9,27), (3,15,27), (6,9,15), (6,9,27), (6,15,27), (9,15,27)

   b) Mean of the sampling distribution of means is
   $\mu_{\bar{x}} = \frac{6+8+9+10+12+13+14+15+16+17}{10} = \frac{120}{10} = 12$

   c) $\sigma^2 = \frac{(6-12)^2+(8-12)^2+(9-12)^2+(10-12)^2+(12-12)^2+(13-12)^2+(14-12)^2+(15-12)^2+(16-12)^2+(17-12)^2}{10}$

   $= 13.3$

   $\therefore \sigma_{\bar{x}} = \sqrt{13.3} = 3.651$

2. A population consist of five numbers 2,3,6,8 and 11. Consider all possible samples of size two which can be drawn with replacement from this population .Find

a) The mean of the population
b) The standard deviation of the population
c) The mean of the sampling distribution of means and
d) The standard deviation of the sampling distribution of means

Sol: a) Mean of the Population is given by

$$\mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$$

b) Variance of the population is given by

$$\sigma^2 = \sum \frac{(x_i - \bar{x})^2}{n}$$

$$= \frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5}$$

$$= \frac{16+9+0+4+25}{5} = 10.8 \qquad \therefore \sigma = 3.29$$

c) Sampling with replacement

The total no.of samples with replacement is $N^n = 5^2 = 25$

∴ List of all possible samples with replacement are

$$\begin{Bmatrix} (2,2),(2,3),(2,6),(2,8),(2,11),(3,2),(3,3)(3,6),(3,8),(3,11) \\ (6,2),(6,3),(6,6),(6,8),(6,11),(8,2),(8,3),(8,6),(8,8),(8,11) \\ (11,2),(11,3),(11,6),(11,8),(11,11) \end{Bmatrix}$$

Now compute the arithmetic mean for each of these 25 samples which gives rise to the distribution of means of the samples known as sampling distribution of means

The samples means are

$$\begin{Bmatrix} 2,2.5,4,5,6.5 \\ 2.5,3,4.5,,5.5,7 \\ 4,4.5,6,7,8.5 \\ 5,5.5,7,8,9.5 \\ 6.5,7,8.5,9.5,11 \end{Bmatrix}$$

And the mean of sampling distribution of means is the mean of these 25 means

$$\mu_{\bar{x}} = \frac{\text{sum of all above sample means}}{25} = \frac{150}{25} = 6$$

d) The variance of the sampling distribution of means is obtained by subtracting the mean 6 from each number in sampling distribution of means and squaring the result ,adding all 25 numbers thus obtained and dividing by 25.

$$\sigma^2 = \frac{(2-6)^2+(2.5-6)^2+(4-6)^2+(5-6)^2+\cdots\ldots(11-6)^2}{25} = \frac{135}{25} = 5.4$$

$$\therefore \sigma = \sqrt{5.4} = 2.32$$

3. When a sample is taken from an infinite population , what happens to the standard error of the mean if the sample size is decreased from 800 to 200

Sol: The standard error of mean = $\frac{\sigma}{\sqrt{n}}$

Sample size = n .let n= n_1=800

Then S.E_1 = $\frac{\sigma}{\sqrt{800}} = \frac{\sigma}{20\sqrt{2}}$

When n_1 is reduced to 200

let n= n_2=200

Then S.E_2 = $\frac{\sigma}{\sqrt{200}} = \frac{\sigma}{10\sqrt{2}}$

$\therefore$ S.E_2 = $\frac{\sigma}{10\sqrt{2}} = 2\left(\frac{\sigma}{20\sqrt{2}}\right) = 2$ (S.E_1)

Hence if sample size is reduced from 800to 200, S. E. of mean will be multiplied by 2

4. The variance of a population is 2 . The size of the sample collected from the population is 169. What is the standard error of mean

Sol: n= The size of the sample =169

σ = S.D of population = $\sqrt{\text{Variance}} = \sqrt{2}$

Standard Error of mean = $\frac{\sigma}{\sqrt{n}} = \frac{\sqrt{2}}{\sqrt{169}} = \frac{1.41}{13} = 0.185$

5. The mean height of students in a college is 155cms and standard deviation is 15 . What is the probability that the mean height of 36 students is less than 157 cms.

Sol: μ = Mean of the population

= Mean height of students of a college = 155cms

n = S.D of population = 15cms

$\bar{x}$ = mean of sample = 157 cms

Now z = $\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{157-155}{\frac{15}{\sqrt{36}}} = \frac{12}{15} = 0.8$

$\therefore$ P ($\bar{x} \leq 157$) = P (z $<$ 0.8) = 0.5 + P (0 $\leq$ z $\leq$ 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. A random sample of size 100 is taken from a population with σ = 5.1 . Given that the sample mean is $\bar{x}$ = 21.6 Construct a 95% confidence limits for the population mean .

Sol: Given $\bar{x}$ = 21.6

$z\alpha_{/2}$ = 1.96, n = 100 , σ = 5.1

$\therefore$ Confidence interval = ($\bar{x} - z\alpha_{/2}.\frac{\sigma}{\sqrt{n}}$, $\bar{x} + z\alpha_{/2}.\frac{\sigma}{\sqrt{n}}$)

$\bar{x} - z\alpha_{/2}.\frac{\sigma}{\sqrt{n}}$ = 21.6 – $\frac{1.96 \times 5.1}{10}$ = 20.6

$\bar{x} + z\alpha_{/2}.\frac{\sigma}{\sqrt{n}}$ = 21.6 + $\frac{1.96 \times 5.1}{10}$ = 22.6

Hence (20.6,22.6) is the confidence interval for the population mean μ

7. It is desired to estimate the mean time of continuous use until an answering machine will first require service . If it can be assumed that σ = 60 days, how large a sample is needed so that one will be able to assert with 90% confidence that the sample mean is off by at most 10 days.

Sol: We have maximum error (E) = 10 days , σ = 60 days and $z_{\alpha/2}$ = 1.645

$\therefore n = \left[\frac{z_{\alpha/2}\cdot\sigma}{E}\right]^2 = \left[\frac{1.645 \times 60}{10}\right]^2 = 97$

8. A random sample of size 64 is taken from a normal population with $\mu = 51.4$ and σ = 6.8.What is the probability that the mean of the sample will a) exceed 52.9 b) fall between 50.5 and 52.3 c) be less than 50.6

Sol: Given n = the size of the sample = 64

μ = the mean of the population = 51.4

σ = the S.D of the population = 6.8

a) P($\bar{x}$ exceed 52.9) = P($\bar{x} > 52.9$)

$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{52.9-51.4}{\frac{6.8}{\sqrt{64}}} = 1.76$

$\therefore P(\bar{x} > 52.9) = P(z > 1.76)$

$= 0.5 - P(0 < z < 1.76)$

$= 0.5 - 0.4608 = 0.0392$

b) P($\bar{x}$ fall between 50.5 and 52.3)

i.e, $P(50.5 < \bar{x} < 52.3) = P(\overline{x_1} < \bar{x} < \overline{x_2})$

$z_1 = \frac{\overline{x_1}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{50.5-51.4}{0.85} = -1.06$

$z_2 = \frac{\overline{x_2}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{52.3-51.4}{0.85} = 1.06$

$P(50.5 < \bar{x} < 52.3) = P(-1.06 < z < 1.06)$

$= P(-1.06 < z < 0) + P(0 < z < 1.06)$

$= P(0 < z < 1.06) + P(0 < z < 1.06)$

$= 2(0.3554) = 0.7108$

c) P($\bar{x}$ will be less than 50.6) = P($\bar{x} < 50.6$)

$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{50.6-51.4}{\frac{6.8}{\sqrt{64}}} = -0.94$

$\therefore P(z < -0.94) = 0.5 - P(0.94 < z < 0)$

$= 0.5 - P(0 < z < 0.94) = 0.50-0.3264$

$= 0.1736$

9. The mean of certain normal population is equal to the standard error of the mean of the samples of 64 from that distribution . Find the probability that the mean of the sample size 36 will be negative.

Sol: The Standard error of mean = $\frac{\sigma}{\sqrt{n}}$

Sample size , n =64

Given mean , μ = Standard error of the mean of the samples

$\mu = \frac{\sigma}{\sqrt{64}} = \frac{\sigma}{8}$

We know z = $\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x}-\frac{\sigma}{8}}{\frac{\sigma}{6}}$

$= \frac{6\bar{x}}{\sigma} - \frac{3}{4}$

If Z < 0.75, $\bar{\bar{x}}$ is negative

P(z < 0.75) = P($-\infty < z < 0.75$)

$= \int_{-\infty}^{0} \phi(z)\, dz + \int_{0}^{0.75} \phi(z) dz = 0.50 + 0.2734$

= 0.7734

10. The guaranteed average life of a certain type of electric bulbs is 1500hrs with a S.D of 10 hrs. It is decided to sample the output so as to ensure that 95% of bulbs do not fall short of the guaranteed average by more than 2% . What will be the minimum sample size ?

Sol : Let n be the size of the sample

The guaranteed mean is 1500

We do not want the mean of the sample to be less than 2% of (1500) i.e, 30 hrs

So 1500 – 30 = 1470

$\therefore \bar{x} > 1470$

$\therefore |z| = \left|\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| = \left|\frac{1470-1500}{\frac{120}{\sqrt{n}}}\right| = \frac{\sqrt{n}}{4}$

From the given condition , the area of the probability normal curve to the left of $\frac{\sqrt{n}}{4}$ should be 0.95

$\therefore$ The area between 0 and $\frac{\sqrt{n}}{4}$ is 0.45

We do not want to know about the bulbs which have life above the guranteed life .

$\therefore \frac{\sqrt{n}}{4}$ = 1.65 i.e., $\sqrt{n}$ = 6.6

$\therefore$ n = 44

11. A normal population has a mean of 0.1 and standard deviation of 2.1 . Find the probability that mean of a sample of size 900 will be negative .

Sol : Given $\mu = 0.1$, $\sigma = 2.1$ and n = 900

The Standard normal variate

Z = $\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x}-\mu}{\frac{2.1}{\sqrt{900}}} = \frac{\bar{x}-0.1}{0.07}$

$\therefore \bar{x}$ = 0.1 + 0.007 z where z $\sim$ N (0 ,1)

$\therefore$ The required probability , that the sample mean is negative is given by

$P(\bar{x} < 0)$ = P (0.1 + 0.07 z < 0)

= P (0.07 z < - 0.1)

= P (z < $\left(\frac{-0.1}{0.07}\right)$

= P (z < -1.43)
= 0.50 – P (0 < z < 1.43)
= 0.50 – 0.4236 = 0.0764

12. In a study of an automobile insurance a random sample of 80 body repair costs had a mean of Rs 472.36 and the S.D of Rs 62.35. If $\bar{\bar{x}}$ is used as a point estimator to the true average repair costs , with what confidence we can assert that the maximum error doesn't exceed Rs 10.

Sol : Size of a random sample , n = 80

The mean of random sample , $\bar{x}$ = Rs 472.36

Standard deviation , σ = Rs 62.35

Maximum error of estimate , E_{max} = Rs 10

We have $E_{max} = Z\alpha_{/2} . \frac{\sigma}{\sqrt{n}}$

i.e., $Z\alpha_{/2} = \frac{E_{max} . \sqrt{n}}{\sigma} = \frac{10\sqrt{80}}{62.35} = \frac{89.4427}{62.35} = 1.4345$

$\therefore Z\alpha_{/2} = 1.43$

The area when z = 1.43 from tables is 0.4236

$\therefore \frac{\alpha}{2} = 0.4236$ i.e ., $\alpha = 0.8472$

$\therefore$ confidence = (1- α) 100% = 84.72 %

Hence we are 84.72% confidence that the maximum error is Rs. 10

13. If we can assert with 95% that the maximum error is 0.05 and P = 0.2 find the size of the sample.

Sol : Given P =0.2 , E = 0.05

We have Q = 0.8 and $Z\alpha_{/2}$ = 1.96 (5% LOS)

We know that maximum error , E = $Z\alpha_{/2} \sqrt{\frac{PQ}{n}}$

$\Rightarrow 0.05 = 1.96 \sqrt{\frac{0.2 \times 0.8}{n}}$

$\Rightarrow$ Sample size , n = $\frac{0.2 \times 0.8 \times (1.96)^2}{(0.05)^2} = 246$

14. The mean and standard deviation of a population are 11,795 and 14,054 respectively . What can one assert with 95 % confidence about the maximum error if $\bar{x}$ = 11,795 and n = 50. And also construct 95% confidence interval for true mean .

Sol: Here mean of population , μ = 11795

S.D of population , σ = 14054

$\bar{x}$ = 11795

$n =$ sample size = 50 , maximum error = $Z\alpha_{/2} . \frac{\sigma}{\sqrt{n}}$

$Z\alpha_{/2}$ for 95% confidence = 1.96

Max. error , $E = Z\alpha_{/2} . \frac{\sigma}{\sqrt{n}} = 1.96 . \frac{14054}{\sqrt{50}} = 3899$

$\therefore$ Confidence interval = ($\bar{x} - Z\alpha_{/2} . \frac{\sigma}{\sqrt{n}} , \bar{x} + Z\alpha_{/2} . \frac{\sigma}{\sqrt{n}}$)

= (11795-3899, 11795+3899)

= (7896, 15694)

15. Find 95% confidence limits for the mean of a normally distributed population from which the following sample was taken 15, 17 , 10 ,18 ,16 ,9, 7, 11, 13 ,14.

Sol: We have $\bar{x} = \frac{15+17+10+18+16+9+7+11+13+14}{10} = 13$

$S^2 = \sum \frac{(x_i - \bar{x})^2}{n-1}$

$= \frac{1}{9}$

$[(15-13)^2 + (15-13)^2 + (15-13)^2 + (15-13)^2 + (15-13)^2 + (15-13)^2 + (15-13)^2 + (15-13)^2 + (15-13)^2 + (15-13)^2]$

$= \frac{40}{3}$

Since $Z\alpha_{/2} = 1.96$, we have

$Z\alpha_{/2} \cdot \frac{s}{\sqrt{n}} = 1.96 \cdot \frac{\sqrt{40}}{\sqrt{10}.\sqrt{3}} = 2.26$

$\therefore$ Confidence limits are $\bar{x} \pm Z\alpha_{/2} \cdot \frac{s}{\sqrt{n}} = 13 \pm 2.26 = (10.74 , 15.26)$

16. A random sample of 100 teachers in a large metropolitan area revealed mean weekly salary of Rs. 487 with a standard deviation Rs.48. With what degree of confidence can we assert that the average weekly of all teachers in the metropolitan area is between 472 to 502 ?

Sol: Given $\mu = 487 , \sigma = 48 , n = 100$

$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

$= \frac{\bar{x} - 487}{\frac{48}{\sqrt{100}}} = \frac{\bar{x} - 487}{4.8}$

Standard variable corresponding to Rs. 472 is

$Z_1 = \frac{472 - 487}{4.8} = -3.125$

Standard vaiable corresponding to Rs. 502

$Z_2 = \frac{502 - 487}{4.8} = 3.125$

Let $\bar{X}$ be the mean salary of teacher . Then

$P(472 < \bar{X} < 502) = P(-3.125 < z < 3.125)$

$= 2(0 < z < 3.125)$

$= 2 \int_0^{3.125} \emptyset(z)\,dz$

$= 2(0.4991) = 0.9982$

Thus we can ascertain with 99.82 % confidence

UNIT-IV

STATISTICAL INFERENCES

Large Samples: Let a random sample of size n $>$30 is defined as large sample.

APPLICATIONS OF LARGE SAMPLES:

TEST OF SIGNIFICANCE OF A SINGLE MEAN

Let a random sample of size n, $\bar{x}$ be the mean of the sample and μ be the population mean.

1. **Null hyopothesis**: H_0: There is no significant difference in the given population mean value say ' μ_0'.

   i.e $H_0: \mu = \mu_0$

2. **Alternative hypothesis**: H_1 :There is some significant difference in the given population mean value.
   i.e
   a)H_1 : μ $\neq \mu_0$ (Two –tailed)
   b) H_1 : μ $> \mu_0$ (Right one tailed)
   c) H_1 : μ $> \mu_0$ (Left one tailed)
3. **Level of significance**: Set the LOS α
4. **Test Statistic**: $z_{cal} = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ (OR) $z_{cal} = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$
5. **Decision /conclusion** : If z_{cal} value $< z_{\alpha}$ value , accept H_0 otherwise reject H_0

CRITICAL VALUES OF Z

LOS ∝	1%	5%	10%
μ≠ μ_0	/Z/>2.58	/Z/>1.96	/Z/>1.645
μ> μ_0	Z>2.33	z>1.645	Z>1.28
μ< μ_0	Z<-2.33	Z<-1.645	Z<-1.28

NOTE: Confidence limits for the mean of the population corresponding to the given sample.

$\mu = \bar{X} \pm Z\alpha_{/2}$ (S.E of $\bar{X}$) i.e,

$\mu = \bar{X} \pm Z\alpha_{/2}\left(\frac{\sigma}{\sqrt{n}}\right)$ (or) $\mu = \bar{X} \pm Z\alpha_{/2}\left(\frac{s}{\sqrt{n}}\right)$

2. TEST OF SIFNIFICANCE FOR DIFFERENCE OF MEANS OF TWO LARGE SAMPLES:

Let $\overline{x_1}$ & $\overline{x_2}$ be the means of the samples of two ramdom sizes n_1 & n_2 drawn from two

populations having means $\mu_1 \& \mu_2$ and SD's $\sigma_1 \& \sigma_2$

i) **Null hyopothesis**: $H_0: \mu_1 = \mu_2$

ii) **Alternative hypothesis** :. a) $H_1 : \mu_1 \neq \mu_2$(Two Tailed)

b) $H_1: \mu_1 < \mu_2$ (Left one tailed)

c) $H_1: \mu_1 > \mu_2$ (Right one tailed)

iii) **Level of Significance**: Set the LOS α

iv) **Test Statistic** : $Z_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{SE\ of(\bar{x}_1 - \bar{x}_2)} = -\frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

Where $\delta = \mu_1 - \mu_2$(where given constant)

Other wise $\delta = \mu_1 - \mu_2$=0

$Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ if $\sigma_1^2 = \sigma_2^2 = \sigma^2$ then $Z_{cal} = \frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

Critical value of Z from normal table at the LOS α

v) **Decision:** If $|Z_{cal}| < Z_{tab}$, accept H_0 otherwise reject H_0

CRITICAL VALUES OF Z

LOS ∝	1%	5%	10%
$\mu \neq \mu_0$	/Z/>2.58	/Z/>1.96	/Z/>1.645
$\mu > \mu_0$	Z>2.33	z>1.645	Z>1.28
$\mu < \mu_0$	Z<-2.33	Z<-1.645	Z<-1.28

NOTE: Confidence limits for difference of means

$$\mu_1 - \mu_2 = (\bar{X}_1 - \bar{X}_2) \pm z_{\alpha/2}[S.E\ of\ (\bar{X}_1 - \bar{X}_2)]$$

$$= (\bar{X}_1 - \bar{X}_2) \pm z_{\alpha/2}\left[\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1}}\right]$$

3. TEST OF SIFNIFICANCE FOR SINGLE PROPORTIONS

Suppose a random sample of size n has a sample proportion p of members possessing a certain attribute (proportion of successes). To test the hypothesis that the proportion P in the population has a specified value P_0.

i) **Null hyopothesis** : $H_0: P = P_0$

ii) **Alternative hypothesis** : a) H_1 : P≠ P_0(Two Tailed test)

b) $H_1 : P < P_0$ (Left one- tailed)

c) H_1 : P $> P_0$ (Right one tailed)

iii) **Test statistic** :$Z_{cal} = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$ when P is the Population proportion $Q = 1 - P$

iv) At specified LOS ∝, critical value of Z

v) **Decision**: If $|z_{cal}| < Z_{tab}$, accept H_0 otherwise reject H_0

CRITICAL VALUES OF Z

LOS ∝	1%	5%	10%
μ≠ μ_0	/Z/>2.58	/Z/>1.96	/Z/>1.645
μ> μ_0	Z>2.33	z>1.645	Z>1.28
μ< μ_0	Z<-2.33	Z<-1.645	Z<-1.28

NOTE : Confidence limits for population proportion

$P = P \pm Z_{\frac{\alpha}{2}(SE\ of\ P)}$

$= P \pm Z_{\frac{\alpha}{2}} \left(\sqrt{\frac{pq}{n}}\right)$

4. TEST FOR EQUALITY OF TWO PROPORTIONS (POPULATIONS)

Let p_1 and p_2 be the sample proportions in two large random samples of sizes n_1 & n_2 drawn from two populations having proportions P_1 & P_2

i) **Null hyopothesis** : $H_0: P_1 = P_2$

ii) **Alternative hypothesis** : a) $H_1 : P_1 \neq P_2$(Two Tailed)

b) $H_1 : P_1 < P_2$ (Left one tailed)

c) $H_1 : P_1 > P_2$ (Right one tailed)

iii) **Test statistic** :$Z_{cal} = \frac{(P_1-P_2)-(P_1-P_2)}{\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_1Q_1}{n_2}}}$ if (P_1-P_2) is given.

If given only sample proportions then

$$Z_{cal} = \frac{p_1-p_2}{\sqrt{\frac{P_1q_1}{n_1}+\frac{P_2q_1}{n_2}}} \text{ where } p_1 = \frac{x_1}{n_1} \text{ \& } p_1 = \frac{x_2}{n_2}$$

OR

$$Z_{cal} = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \text{ Where } p = \frac{n_1p_1+n_2p_2}{n_1+n_2} = \frac{x_1+x_2}{n_1+n_2} \text{ and } q = 1- p$$

iv) At specified LOS $\propto$ critical value of 'Z'

v) **Decision**: If $|Z_{cal}| < Z_{Tab}$, accept H_0 otherwise reject H_0

CRITICAL VALUES OF Z

LOS $\propto$	1%	5%	10%
$\mu \neq \mu_0$	/Z/>2.58	/Z/>1.96	/Z/>1.645
$\mu > \mu_0$	Z>2.33	z>1.645	Z>1.28
$\mu < \mu_0$	Z<-2.33	Z<-1.645	Z<-1.28

NOTE: Confidence limits for difference of population proportions

$$P_1 - P_2 = (p_1 - p_2) \pm Z_{\frac{\propto}{2}} (S.E\ of\ P_1 - P_2)$$

Problems:

1. A sample of 64 students have a mean weight of 70 kgs . Can this be regarded as asample mean from a population with mean weight 56 kgs and standard deviation 25 kgs.
   Sol : Given $\bar{x}$ = mean of he sample = 70 kgs
   μ = Mean of the population = 56 kgs
   σ = S.D of population = 25 kgs

and n = Sample size = 64

i) Null Hypothesis H_0 : A Sample of 64 students with mean weight 70 kgs be regarded as a sample from a population with mean weight 56 kgs and standard deviation 25 kgs. i.e., $H_0 : \mu =$ 70 kgs

ii) Alternative Hypothesis H_1 : Sample cannot be regarded as one coming from the population . i.e., $H_1 : \mu \neq$ 70 kgs (Two –tailed test)

iii) Level of significance : $\alpha = 0.05$ (Z_α = 1.96)

iv) Test Statistic : $Z_{cal} = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{70-56}{\frac{25}{\sqrt{64}}} = 4.48$

v) Conclusion: Since $|Z_{cal}|\ value > Z_\alpha$ value , we reject H_0

∴ Sample cannot be regarded as one coming from the population

2. In a random sample of 60 workers , the average time taken by them to get to work is 33.8 minutes with a standard deviation of 6.1 minutes . Can we reject the null hypothesis μ = 32.6 in favor of alternative null hypothesis μ > 32.6 at α = 0.05 LOS

Sol : Given n = 60 , $\bar{x}$ = 33.8 , μ = 32.6 and σ = 6.1

i) Null Hypothesis $H_0 : \mu = 32.6$

ii) Alternative Hypothesis $H_1 : \mu > 32.6$ (Right one tailed test)

iii) Level of significance : $\alpha = 0.01$ (Z_α = 2.33)

iv) Test Statistic : $Z_{cal} = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{33.8-32.6}{\frac{6.1}{\sqrt{60}}} = \frac{1.2}{0.7875} = 1.5238$

v) Conclusion: Since $Z_{cal}\ value < Z_\alpha$ value , we accept H_0

3. A sample of 400 items is taken from a population whose standard deviation is 10 . The mean of the sample is 40 . Test whether the sample has come from a population with mean 38 . Also calculate 95% confidence limits for the population .

Sol : Given n = 400 , $\bar{x}$ = 40 , μ = 38 and σ = 10

i) Null Hypothesis $H_0 : \mu = 38$

ii) Alternative Hypothesis $H_1 : \mu \neq 38$ (Two –tailed test)

iii) Level of significance : $\alpha = 0.05$ (Z_α = 1.96)

iv) Test Statistic : $Z_{cal} = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{38-40}{\frac{10}{\sqrt{400}}} = \frac{-2}{0.5} = -4$

v) Conclusion: Since $|Z_{cal}|\ value > Z_\alpha$ value , we reject H_0

i.e., the sample is not from the population whose is 38.

∴ 95% confidence interval is $\left(\bar{x} - 1.96.\frac{\sigma}{\sqrt{n}} , \bar{x} + 1.96.\frac{\sigma}{\sqrt{n}}\right)$

i.e., $\left(40 - \frac{1.96(10)}{\sqrt{400}} , 40 + \frac{1.96(10)}{\sqrt{400}}\right)$

$= \left(40 - \frac{1.96(10)}{20} , 40 + \frac{1.96(10)}{20}\right)$

= (40 – 0.98 , 40 + 0.98)

= (39.02 , 40.98)

4. An insurance agent has claimed that the average age of policy holders who issue through him is less than the average for all agents which is 30.5. A random sample of 100 policy holders who had issued through him gave the following age distribution .

Age	16-20	21-25	26-30	31-35	36-40
No# of persons	12	22	20	30	16

Calculate the arithmetic mean and standard deviation of this distribution and use these values to test his claim at 5% los.

Sol : Take A = 28 where A – Assumed mean

$d_i = x_i$ – A

$\bar{x}$ = A + $\frac{h \sum f_i d_i}{N}$

= 28 + $\frac{5 \times 16}{100}$ = 28.8

S.D : S = h $\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$ = 5. $\sqrt{\frac{164}{100} - \left(\frac{16}{100}\right)^2}$ = 6.35

i) Null Hypothesis H_0 : The sample is drawn from population with mean μ

ii) i.e., $H_0 : \mu$ = 30.5 years

iii) Alternative Hypothesis $H_1 : \mu < 30.5$ (Left one –tailed test)

iv) Level of significance : $\alpha = 0.05$ (Z_α = 1.645)

v) Test Statistic : $Z_{cal} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{28.8 - 30.5}{\frac{6.35}{\sqrt{100}}} = -2.677$

vi) Conclusion: Since $|Z_{cal}|\ value > Z_\alpha$ value , we reject H_0

i.e., the sample is not drawn from the population with μ = 30.5 years .

5. An ambulance service claims that it takes on the average less than 10 minutes to reach its destination in emergency calls . A sample of 36 calls has a mean of 11 minutes and the variance of 16 minutes .Test the claim at 0.05 los?

Sol : Given n = 36 , $\bar{x}$ =11 , μ = 10 and $\sigma = \sqrt{16} = 4$

i) Null Hypothesis $H_0 : \mu$ = 10

ii) Alternative Hypothesis $H_1 : \mu < 10$ (Left one –tailed test)

iii) Level of significance : $\alpha = 0.05$ (Z_α = 1.645)

iv) Test Statistic : $Z_{cal} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{11 - 10}{\frac{4}{\sqrt{36}}} = \frac{6}{4} = 1.5$

v) Conclusion: Since $|Z_{cal}|\ value < Z_\alpha$ value , we accept H_0

6. The means of two large samples of sizes 1000 and 2000 members are 67.5 inches and 68 inches respectively . Can the samples be regarded as drawn from the same population of S.D 2.5 inches.

Sol: Let μ_1 and μ_2 be the means of the two populations

Given n_1 = 1000 , n_2 = 2000 and $\bar{x}_1$ = 67.5 inches , $\bar{x}_2$ = 68 inches

Population S.D, σ = 2.5 inches

i) Null Hypothesis H_0 :The samples have been drawn from the same population of S.D 2.5 inches

i.e., $H_0 : \mu_1 = \mu_2$

ii) Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$ (Two – Tailed test)

iii) Level of significance : $\alpha = 0.05$ (Z_α = 1.96)

iv) Test Statistic : $Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{67.5-68}{\sqrt{(2.5)^2\left(\frac{1}{1000}+\frac{1}{2000}\right)}} = \frac{-0.5}{0.0968} = -5.16$

v) Conclusion: Since $|Z_{cal}|\ value > Z_\alpha$ value , we reject H_0

Hence , we conclude that the samples are not drawn from the same population of S.D 2.5 inches.

7. Samples of students were drawn from two universities and from their weights in kilograms , mean and standard deviations are calculated and shown below. Make a large sample test to test the significance of the difference between the means.

	Mean	S .D	Size of the sample
University A	55	10	400
University B	57	15	100

Sol: : Let μ_1 and μ_2 be the means of the two populations

Given n_1 = 400 , n_2 = 100 and $\bar{x}_1$ = 55 kgs , $\bar{x}_2$ = 57 kgs

σ_1 = 10 and σ_2 = 15

i) Null Hypothesis $H_0 : \mu_1 = \mu_2$

ii) Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$ (Two – Tailed test)

iii) Level of significance : $\alpha = 0.05$ (Z_α = 1.96)

iv) Test Statistic : $Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} = \frac{55-57}{\sqrt{\frac{10^2}{400}+\frac{15^2}{100}}} = \frac{-2}{\sqrt{\frac{1}{4}+\frac{9}{4}}} = -1.26$

v) Conclusion: Since $|Z_{cal}|\ value < Z_\alpha$ value , we accept H_0

Hence , we conclude that there is no significant difference between the means

8. The average marks scored by 32 boys is 72 with a S.D of 8 . While that for 36 girls is 70 with a S.D of 6. Does this data indicate that the boys perform better than girls at 5% los ?

Sol: Let μ_1 and μ_2 be the means of the two populations

Given n_1 = 32 , n_2 = 36 and $\bar{x}_1$ = 72 , $\bar{x}_2$ = 70

σ_1 = 8 and σ_2 = 6

i) Null Hypothesis $H_0 : \mu_1 = \mu_2$

ii) Alternative Hypothesis $H_1 : \mu_1 > \mu_2$ (Right One Tailed test)

iii) Level of significance : $\alpha = 0.05$ (Z_α = 1.645)

iv) Test Statistic : $Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} = \frac{72-70}{\sqrt{\frac{8^2}{32}+\frac{6^2}{36}}} = \frac{2}{\sqrt{2+1}} = 1.1547$

v) Conclusion: Since $|Z_{cal}|\ value < Z_\alpha$ value , we accept H_0

Hence , we conclude that the performance of boys and girls is the same

9. A sample of the height of 6400 Englishmen has a mean of 67.85 inches and a S.D of 2.56 inches while another sample of heights of 1600 Austrians has a mean of 68.55 inches and S.D of 2.52 inches. Do the data indicate that Austrians are on the average taller than the Englishmen ? (Use $\alpha\ as\ 0.01$)

Sol : Let μ_1 and μ_2 be the means of the two populations

Given n_1 = 6400 , n_2 = 1600 and $\bar{x}_1$ = 67.85 , $\bar{x}_2$ = 68.55

σ_1 = 2.56 and σ_2 = 2.52

i) Null Hypothesis $H_0 : \mu_1 = \mu_2$

ii) Alternative Hypothesis $H_1 : \mu_1 < \mu_2$ (Left One Tailed test)

iii) Level of significance : $\alpha = 0.01$ (Z_α = - 2.33)

iv) Test Statistic : $Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{67.85 - 68.55}{\sqrt{\frac{2.56^2}{6400} + \frac{2.52^2}{1600}}}$

$$= \frac{67.85 - 68.55}{\sqrt{\frac{6.5536}{6400} + \frac{6.35}{1600}}}$$

$$= \frac{-0.7}{\sqrt{0.001 + 0.004}} = \frac{-0.7}{0.0707} = -9.9$$

v) Conclusion: Since $|Z_{cal}|\, value > Z_\alpha$ value , we reject H_0

Hence , we conclude that Australians are taller than Englishmen.

10. At a certain large university a sociologist speculates that male students spend considerably more money on junk food than female students. To test her hypothesis the sociologist randomly selects from records the names of 200 students . Of thee , 125 are men and 75 are women . The mean of the average amount spent on junk food per week by the men is Rs. 400 and S.D is 100. For the women the sample mean is Rs. 450 and S.D is 150. Test the hypothesis at 5 % los ?

Sol: Let μ_1 and μ_2 be the means of the two populations

Given n_1 = 125 , n_2 = 75 and $\bar{x}_1$ = Mean of men = 400 , $\bar{x}_2$ = Mean of women = 450

σ_1 = 100 and σ_2 = 150

i) Null Hypothesis $H_0 : \mu_1 = \mu_2$

ii) Alternative Hypothesis $H_1 : \mu_1 > \mu_2$ (Right One Tailed test)

iii) Level of significance : $\alpha = 0.05$ (Z_α = 1.645)

iv) Test Statistic : $Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{400 - 450}{\sqrt{\frac{100^2}{125} + \frac{150^2}{75}}}$

$$= \frac{-50}{\sqrt{80 + 300}}$$

$$= \frac{-50}{\sqrt{380}} = \frac{-50}{19.49} = -2.5654$$

v) Conclusion: Since $Z_{cal} value < Z_\alpha$ value , we accept H_0

Hence , we conclude that difference between the means are equal

11. The research investigator is interested in studying whether there is a significant difference in the salaries of MBA grads in two cities. A random sample of size 100 from city A yields an average income of Rs. 20,150 . Another random sample of size 60 from city B yields an average income of Rs. 20,250. If the variance are given as σ_1^2 = 40,000 and σ_2^2 = 32,400 respectively . Test the equality of means and also construct 95% confidence limits.

Sol: Let μ_1 and μ_2 be the means of the two populations

Given n_1 = 100 , n_2 = 60 and $\bar{x}_1$ = Mean of city A = 20,150 , $\bar{x}_2$ = Mean of city B = 20,250

${\sigma_1}^2$ = 40,000 and ${\sigma_2}^2$ = 32,400

i) Null Hypothesis $H_0 : \mu_1 = \mu_2$

ii) Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$ (Two -Tailed test)

iii) Level of significance : $\alpha = 0.05$ (Z_α = 1.96)

iv) Test Statistic : $Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}} = \frac{20{,}150 - 20{,}250}{\sqrt{\frac{40000}{100} + \frac{32400}{60}}}$

$= \frac{100}{\sqrt{400 + 540}}$

$= \frac{100}{30.66}$ = 3.26

v) Conclusion: Since $Z_{cal} value > Z_\alpha$ value , we reject H_0

Hence , we conclude that there is a significant difference in the salaries of MBA grades two cities.

$\therefore$ 95% confidence interval is $\mu_1 - \mu_2 = (\overline{x_1} - \overline{x_2}) \pm 1.96 \sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}$

= (20,150 – 20,250)) $\pm$ 1.96 $\sqrt{\frac{40000}{100} + \frac{32400}{60}}$ = (39.90, 160.09)

12. A die was thrown 9000 times and of these 3220 yielded a 3 or 4. Is this consistent with the hypothesis that the die was unbiased?

Sol : Given n = 9000

P = Population of proportion of successes

= P(getting a 3 or 4) = $\frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$ 0.3333

Q = 1- P = 0.6667

P = Proportion of successes of getting 3 or 4 in 9000 times = $\frac{3220}{9000}$ = 0.3578

i) Null Hypothesis H_0 : The die is unbiased

i.e., H_0 : P = 0.33

ii) Alternative Hypothesis H_1 : The die is biased

i.e., H_1 : P $\neq$ 0.33 (Two –Tailed test)

iii) Level of Significance : $\alpha = 0.05$ (Z_α = 1.96)

iv) Test Statistic : $Z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.3578 - 0.3333}{\sqrt{\frac{(0.3333)(0.6667)}{9000}}}$ = 4.94

v)Conclusion: Since $Z_{cal} value > Z_\alpha$ value , we reject H_0

Hence , we conclude that the die is biased.

13. In a random sample of 125 cool drinkers , 68 said they prefer thumsup to Pepsi . Test the null hypothesis P = 0.5 against the alternative hypothesis hypothesis P > 0.5?

Sol : Given n = 125 , x = 68 and p = $\frac{x}{n} = \frac{68}{125}$ = 0.544

i) Null Hypothesis H_0 : P = 0.5

ii) Alternative Hypothesis H_1 : P > 0.5(Right One Tailed test)

iii) Level of Significance : $\alpha = 0.05$ (Z_α = 1.645)

iv) Test Statistic : $Z_{cal} = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.544-0.5}{\sqrt{\frac{(0.5)(0.5)}{125}}} = 0.9839$

v) Conclusion: Since $Z_{cal} value < Z_\alpha$ value , we accept H_0

14. A manufacturer claimed that at least 95% of the equipment which he supplied to a factory conformed to specifications . An experiment of a sample of 200 piece of equipment revealed that 18 were faulty .Test the claim at 5% los ?

Sol : Given n = 200

Number of pieces confirming to specifications = 200-18 = 182

∴ p = Proportion of pieces confirming to specification = $\frac{182}{200}$ = 0.91

P = Population proportion = $\frac{95}{100}$ = 0.95

i) Null Hypothesis H_0 : P = 0.95

ii) Alternative Hypothesis H_1 : P < 0.95(Left One Tailed test)

iii) Level of Significance : $\alpha = 0.05$ (Z_α = -1.645)

iv) Test Statistic : $Z_{cal} = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.91-0.95}{\sqrt{\frac{0.95 \times 0.05}{200}}} = -2.59$

v) Conclusion: We reject H_0

Hence , we conclude that the manufacture's claim is rejected.

15. Among 900 people in a state 90 are found to be chapatti eaters . Construct 99% confidence interval for the true proportion and also test the hypothesis for single proportion ?

Sol: Given x = 90 , n = 900

∴ p = $\frac{x}{n} = \frac{90}{100} = \frac{1}{10} = 0.1$

And q = 1- p= 0.9

Now $\sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.1)(0.9)}{900}} = 0.01$

Confidence interval is $P = p \pm Z_{\frac{\alpha}{2}} \left(\sqrt{\frac{pq}{n}}\right)$

i.e., (0.1- 0.03 , 0.1 + 0.03)

= (0.07 , 0.13)

i) Null Hypothesis H_0 : P = 0.5

ii) Alternative Hypothesis H_1 : P ≠0.5(Two Tailed test)

iii) Level of Significance : $\alpha = 0.01$ (Z_α = 2.58)

iv) Test Statistic : $Z_{cal} = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.1-0.5}{\sqrt{\frac{0.5 \times 0.5}{900}}} = -24.39$

v) Conclusion: Since $|Z_{cal}| value > Z_\alpha$ value , we reject H_0

16. Random samples of 400 men and 200 women in a locality were asked whether they would like to have a bus stop a bus stop near their residence . 200 men and 40 women in favor of the proposal . Test the significance between the difference of two proportions at 5% los ?

Sol: Let P_1 and P_2 be the population proportions in a locality who favor the bus stop

Given n_1 = Number of men = 400

n_2 = number of women = 200

x_1 = Number of men in favor of the bus stop = 200

x_2 = Number of women in favor of the bus stop 40

$\therefore p_1 = \frac{x_1}{n_1} = \frac{200}{400} = \frac{1}{2}$ and $p_2 = \frac{x_2}{n_2} = \frac{40}{200} = \frac{1}{5}$

i) Null Hypothesis $H_0 : P_1 = P_2$

ii) Alternative Hypothesis $H_1 : P_1 \neq P_2$ (Two Tailed test)

iii) Level of Significance : $\alpha = 0.05$ (Z_α = 1.96)

iv) Test Statistic : $Z_{cal} = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

We have p = $\frac{n_1p_1+n_2p_2}{n_1+n_2} = \frac{x_1+x_2}{n_1+n_2} = \frac{200+40}{400+200} = \frac{240}{600} = \frac{2}{5}$

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

$= \frac{0.5-0.2}{\sqrt{(0.4)(0.6)\left(\frac{1}{400}+\frac{1}{200}\right)}} = 7.07$

v) Conclusion: Since $|Z_{cal}| value > Z_\alpha$ value , we reject H_0

Hence we conclude that there is difference between the men and women in their attitude towards the bus stop near their residence.

17. A machine puts out 16 imperfect articles in a sample of 500 articles . After the machine is overhauled it puts out 3 imperfect articles in a sample of 100 articles . Has the machine is improved ?

Sol : Let P_1 and P_2 be the proportions of imperfect articles in the proportion of articles manufactured by the machine before and after overhauling , respectively.

Given n_1 = Sample size before the machine overhauling = 500

n_2 = Sample size after the machine overhauling = 100

x_1 = Number of imperfect articles before overhauling = 16

x_2 = Number of imperfect articles after overhauling = 3

$\therefore p_1 = \frac{x_1}{n_1} = \frac{16}{500} = 0.032$ and $p_2 = \frac{x_2}{n_2} = \frac{3}{100} = 0.03$

i) Null Hypothesis $H_0 : P_1 = P_2$

ii) Alternative Hypothesis $H_1 : P_1 > P_2$ (Left one Tailed test)

iii) Level of Significance : $\alpha = 0.05$ (Z_α = 1.645)

iv) Test Statistic : $Z_{cal} = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

We have p = $\frac{n_1p_1+n_2p_2}{n_1+n_2} = \frac{x_1+x_2}{n_1+n_2} = \frac{16+3}{500+100} = \frac{19}{600}$ = 0.032

q = 1- p = 0.968

$= \frac{0.032-0.03}{\sqrt{(0.032)(0.968)\left(\frac{1}{500}+\frac{1}{100}\right)}}$

$\frac{0.002}{0.019}$ = 0.104

v) Conclusion: Since $|Z_{cal}|value < Z_\alpha$ value , we accept H_0

Hence we conclude that the machine has improved.

18. In an investigation on the machine performance the following results are obtained .

	No# of units inspected	No# of defectives
Machine 1	375	17
Machine 2	450	22

Test whether there is any significant performance of two machines at α = 0.05

Sol: Let P_1 and P_2 be the proportions of defective units in the population of units inspected in machine 1 and Machine 2 respectively.

Given n_1 = Sample size of the Machine 1 = 375

n_2 = Sample size of the Machine 2 = 450

x_1 = Number of defectives of the Machine 1 = 17

x_2 = Number of defectives of the Machine 2 = 22

$\therefore p_1 = \frac{x_1}{n_1} = \frac{17}{375}$ = 0.045 and $p_2 = \frac{x_2}{n_2} = \frac{22}{450}$ = 0.049

i) Null Hypothesis $H_0 : P_1 = P_2$

ii) Alternative Hypothesis $H_1 : P_1 \neq P_2$ (Two Tailed test)

iii) Level of Significance : $\alpha = 0.05$ (Z_α = 1.96)

iv) Test Statistic : $Z_{cal} = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

We have p = $\frac{n_1p_1+n_2p_2}{n_1+n_2} = \frac{x_1+x_2}{n_1+n_2} = \frac{17+22}{375+450} = \frac{39}{825}$ = 0.047

q = 1- p = 1- 0.047 = 0.953

$= \frac{0.045-0.049}{\sqrt{(0.047)(0.953)\left(\frac{1}{375}+\frac{1}{450}\right)}}$

= - 0.267

v) Conclusion: Since $|Z_{cal}|value < Z_\alpha$ value , we accept H_0

Hence we conclude that there is no significant difference in performance of machines.

19. A cigarette manufacturing firm claims that its brand A line of cigarettes outsells its brand B by 8% . If it is found that 42 out of 200 smokers prefer brand A and 18 out of another sample of 100 smokers prefer brand B . Test whether 8% difference is a valid claim?

Sol: Given n_1 = 200

n_2 = 100

x_1 = Number of smokers preferring brand A= 42

x_2 = Number of smokers preferring brand B = 18

$\therefore p_1 = \frac{x_1}{n_1} = \frac{42}{200} = 0.21$ and $p_2 = \frac{x_2}{n_2} = \frac{18}{100} = 0.18$

and $P_1 - P_2 = 8\% = 0.08$

i) Null Hypothesis $H_0 : P_1 - P_2 = 0.08$

ii) Alternative Hypothesis $H_1 : P_1 - P_2 \neq 0.08$ (Two Tailed test)

iii) Level of Significance : $\alpha = 0.05$ ($Z_\alpha = 1.96$)

iv) Test Statistic : $Z_{cal} = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

We have p = $\frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{42+18}{200+100} = \frac{60}{300} = 0.2$

q = 1- p = 1- 0.2 = 0.8

$Z_{cal} = \frac{(0.21 - 0.18) - 0.08}{\sqrt{(0.2)(0.8)\left(\frac{1}{200} + \frac{1}{100}\right)}}$

$= \frac{-0.05}{0.0489} = -1.02$

v) Conclusion: Since $|Z_{cal}|value < Z_\alpha$ value , we accept H_0

Hence we conclude that 8% difference in the sale of two brands of cigarettes is a valid claim.

20. In a city A , 20% of a random sample of 900 schoolboys has a certain slight physical defect . In another city B ,18.5% of a random sample of 1600 school boys has the same defect . Is the difference between the proportions significant at 5% los?

Sol: Given n_1 = 900

n_2 = 1600

x_1 = 20% of 900 = 180

x_2 = 18.5% of 1600 = 296

$\therefore p_1 = \frac{x_1}{n_1} = \frac{180}{900} = 0.2$ and $p_2 = \frac{x_2}{n_2} = \frac{296}{1600} = 0.185$

i) Null Hypothesis $H_0 : P_1 = P_2$

ii) Alternative Hypothesis $H_1 : P_1 \neq P_2$ (Two Tailed test)

iii) Level of Significance : $\alpha = 0.05$ ($Z_\alpha = 1.96$)

iv) Test Statistic : $Z_{cal} = \frac{(p_1 - p_2)}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

We have p = $\frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{180+296}{900+1600} = \frac{476}{2500} = 0.19$

q = 1- p = 1- 0.19 = 0.81

$$Z_{cal} = \frac{0.2 - 0.185}{\sqrt{(0.19)(0.81)\left(\frac{1}{900} + \frac{1}{1600}\right)}}$$

$$= \frac{-0.015}{0.01634} = -0.918$$

v) Conclusion: Since $|Z_{cal}| value < Z_{\alpha}$ value , we accept H_0

Hence we conclude that there is no significant difference between the proportions.

SMALL SAMPLES

INTRODUCTION: When the sample size n < 30, then if is referred to as small samples. In this sampling distribution in many cases may not be normal ie., we will not be justified in estimating the population parameters as equal to the corresponding sample values.

DEGREE OF FREEDOM: The number of independent variates which make up the statistic is known as the degrees of freedom (d.f) and it is denoted by ϑ.

FOR EXAMPLE: If $x_1 + x_2 + x_3 = 50$ and we assign any values to two os the variables (say x_1, x_2), then the values of x_3 will be known. Thus, the two variables are free and independent choices for finding the third.

In general, the number of degrees of freedom is equal to the total number of observations less the number of independent constraints imposed on the observations.

FOR EXAMPLE, in a set of data of n observations, if K is the number of independent constraints then $\vartheta = n - k$

STUDENT'S t-DISTRIBUTION OR t-DISTRIBUTION:

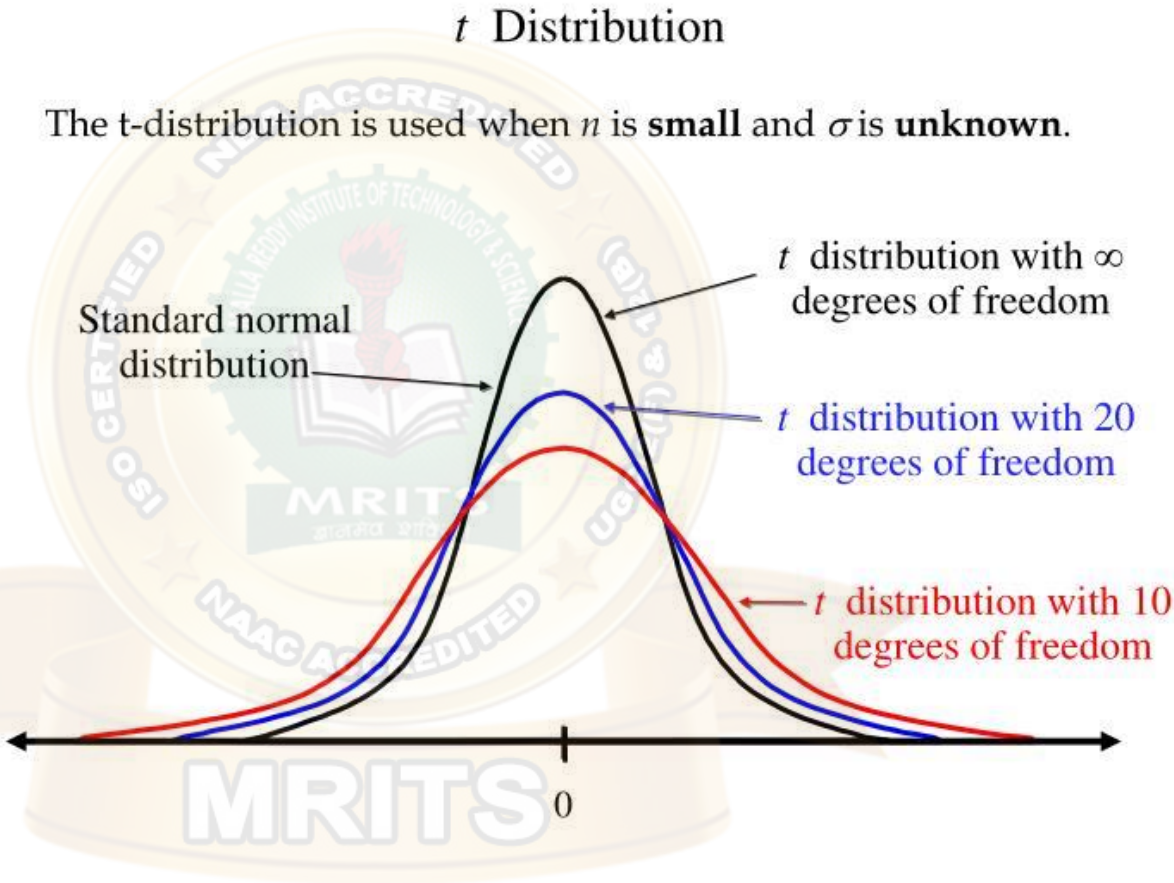
Let $\bar{X}$ be the mean of a random sample of size n, taken from a normal population having the mean μ and the variance σ^2, and sample variance $S^2 = \sum \frac{(x_i - \bar{x})^2}{n-1}$, then

$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$ is a random variable having the $t-$ distribution with $\vartheta = n - 1$ degrees of freedom.

PROPERTIES OF $t-$ DISTRIBUTION:

1. The shape of $t-$distribution is bell shaped, which is similar to that of normal distribution and is symmetrical about the mean.

2. The mean of the standard normal distribution as well as t −distribution is zero, but the variance of t −distrubution depends upon the parometer ϑ which is called the degrees of freedom.
3. The variance of t −distribution exceeds 1, but approaches 1 as $n \to \infty$.



APPLICATIONS OF t – DISTRIBUTIONS:

1. **To test the significance of the sample mean, When population variance is not given:**
   Let $\bar{x}$ be the mean of the sample and n be the size of the sample 'σ' be the standard deviation of the population and μ be the mean of the population.
   Then the student t − distribution is defined by the statistic
   $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n-1}}}$ if s is given directly

   If 'σ' is unknown, then $t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$ where

$S^2 = \sum \frac{(X_i - \bar{X})^2}{n-1}$

Note : Confidence limits for mean $\mu = \bar{x} \pm t_\alpha(S/\sqrt{n})$ or $\mu = \bar{x} \pm t_\alpha(S/\sqrt{n-1})$

2. To test the significance of the difference between means of the two independent samples :

To test the significant difference between the sample means $\bar{x}_1$ and $\bar{x}_2$ of two independent samples of sizes n_1 and n_2, with the same variance .

We use statistic

$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$-------(1) where

$\bar{x}_1 = \frac{\sum x_1}{n_1}, \bar{x}_2 = \frac{\sum x_2}{n_2}$ and

$S^2 = \frac{1}{n_1 + n_2 - 2} [\sum(x_1 - \overline{x_1})^2 + \sum(x_2 - \overline{x_2})^2$

OR $S^2 = \frac{1}{n_1 + n_2 - 2}[(n_1 s_1^2) + (n_2 s_2^2)]$

Where s_1 and s_2 are sample standard deviations.

Note: Confidence limits for difference of means : $\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm t_\alpha\left(\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$

Paired t- test (Test the significance of the difference between means of two dependent samples) :

Paired observations arise in many practical situations where each homogenous experimental unit receives both population condition.

FOR EXAMPLE: To test the effectiveness of 'drug' some // person's blood pressure is measured before and after the intake of certain drug. Here the individual person is the experimental unit and the two populations are blood pressure "before" and "after" the drug is given

Paired t-test is applied for n paired observations by taking the differences d_1, d_2 ------d_n of the paired data. To test whether the differences d_i from a random sample of a population with mean μ.

$t = \frac{\bar{d}}{s/\sqrt{n}}$ *where* $\bar{d} = \frac{1}{n}\epsilon\, d_i$ and $s^2 = \frac{1}{n-1}\sum(d - \bar{d})^2$

Problems:

1. A sample of 26 bulbs gives a mean life of 990 hours with a S.D of 20 hours. The manufacturer claims that the mean life of bulbs is 1000 hours . Is the sample not upto the standard?

   Sol: Given n = 26

   $\bar{x}$ = 990

   μ = 1000 and S.D i.e., s = 20

   i) **Null Hypothesis** : H_0 : μ = 1000

   ii) **Alternative Hypothesis:** H_1 : μ < 1000(Left one tailed test)

   (Since it is given below standard)

   iii) **Level of significance** : $\alpha = 0.05$

   t tabulated value with 25 degrees of freedom for left tailed test is 1.708

   iv) **Test Statistic** : $t_{cal} = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n-1}}} = \frac{990-1000}{\frac{20}{\sqrt{25}}} = -2.5$

   v) **Conclusion:** Since $|t_{cal}|$ value > t_α value , we reject H_0

   Hence we conclude that the sample is not upto the standard.

2. A random sample of size 16 values from a normal population showed a mean of 53 and sum of squares of deviations from the mean equals to 150 . Can this sample be regarded as taken from the population having 56 as mean ? Obtain 95% confidence limits of the mean of the population.?

   Sol: a) Given n = 16

   $\bar{x}$ = 53

   μ = 56 and $\sum(x_i - \bar{x})^2$ = 150

$\therefore S^2 = \frac{\sum(x_i-\bar{x})^2}{n-1} = \frac{150}{15} = 10 \Rightarrow S = \sqrt{10}$

Degrees of freedom ϑ = n-1 = 16-1 =15

   i) **Null Hypothesis** H_0 : μ = 56

   ii) **Alternative Hypothesis** H_1 : $\mu \neq 56$ (Two tailed test)

   iii) **Level of significance** : $\alpha = 0.05$

   t tabulated value with 15 degrees of freedom for two tailed test is 2.13

   iv) **Test Statistic** : $t_{cal} = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}} = \frac{53-56}{\frac{\sqrt{10}}{\sqrt{15}}} = -3.79$

   v) **Conclusion**: Since $|t_{cal}|$ *value* > t_α value , we reject H_0

   Hence we conclude that the sample cannot be regarded as taken from population.

   b) The 95% confidence limits of the mean of the population are given by

   $\bar{x} \pm t_{0.05}\frac{S}{\sqrt{n}} = 53 \pm 2.13 \times 0.79$

   $= 53 \pm 1.6827$

   = 54.68 and 51.31

   $\therefore$ 95% confidence limits are(51.31, 54.68)

3. A random sample of 10 boys had the following I.Q's : 70, 120 ,110, 101,88, 83,95,98,107 and 100.
   a) Do these data support the assumption of a population mean I.Q of 100?
   b) Find a reasonable range in which most of the mean I.Q values of samples of 10 boys lie

Sol: Since mean and s.d are not given

We have to determine these

x	$x-\bar{x}$	$(x-\bar{x})^2$
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
$\sum x = 972$		$\sum (x-\bar{x})^2 = 1833.60$

Mean , $\bar{x} = \frac{\sum x}{n} = \frac{972}{10} = 97.2$ and

$S^2 = \frac{1}{n-1}\sum(x-\bar{x})^2 = \frac{1833.6}{9}$

$\therefore S = \sqrt{203.73} = 14.27$

i) **Null Hypothesis H_0 :** $\mu = 100$

ii) **Alternative Hypothesis H_1 :** $\mu \neq 100$ (Two tailed test)

iii) **Level of significance** : $\alpha = 0.05$

t tabulated value with 9 degrees of freedom for two tailed test is 2.26

iv) **Test Statistic** : $t_{cal} = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}} = \frac{97.2-100}{\frac{14.27}{\sqrt{10}}} = -0.62$

v) **Conclusion**: Since $|t_{cal}|\ value < t_\alpha$ value , we accept H_0

Hence we conclude that the data support the assumption of mean I.Q of 100 in the population.

b) The 95% confidence limits of the mean of the population are given by

$\bar{x} \pm t_{0.05} \frac{S}{\sqrt{n}} = 97.2 \pm 2.26 \times 4.512$

$= 97.2 \pm 10.198$

$= 107.4$ and 87

$\therefore$ 95% confidence limits are(87, 107.4)

4. Samples of two types of electric bulbs were tested for length of life and following data were obtained

Type 1	Type 2
Sample number ,n_1 = 8	n_2 = 7
Sample mean , $\overline{x_1}$ = 1234	$\overline{x_2}$ = 1036
Sample S.D , s_1= 36	s_2= 40

Is the difference in the mean sufficient to warrant that type 1 is superior to type 2 regarding length of life .

Sol: i) **Null Hypothesis** H_0 : The two types of electric bulbs are identical

i.e., H_0: $\mu_1 = \mu_2$

ii) **Alternative Hypothesis** H_1 : $\mu_1 \neq \mu_2$

iii)**Test Statistic** : $t_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

Where $S^2 = \frac{n_1 s_1^{\,2} + n_1 s_1^{\,2}}{n_1 + n_2}$

$= \frac{1}{8+7-2}(8(36)^2 + 7(40)^2) = 1659.08$

$\therefore t = \frac{1234 - 1036}{\sqrt{1659.08\left(\frac{1}{8}+\frac{1}{7}\right)}} = 9.39$

iv)**Degrees of freedom** = 8+7-2 =13 ,tabulated value of t for 13 d.f at 5% los is 2.16

v)**Conclusion**: Since $|t_{cal}|\ value > t_\alpha$ value , we reject H_0

Hence we conclude that the two types 1 and 2 of electric bulbs are not identical .

5. Two horses A and B were tested according to the time to run a particular track with the following results .

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29	

Test whether the two horses have the same running capacity

Sol: Given n_1 = 7 , n_2 = 6

We first compute the sample means and standard deviations

$\bar{x}$ = Mean of the first sample = $\frac{1}{7}$ (28 + 30 + 32 + 33 + 33 + 29 + 34)

$= \frac{1}{7}(219) = 31.286$

$\bar{y}$ = Mean of the second sample = $\frac{1}{6}$ (29 + 30 + 30 + 24 + 27 + 29)

$= \frac{1}{6}(169) = 28.16$

x	$x - \bar{x}$	$(x - \bar{x})^2$	y	$y - \bar{y}$	$(y - \bar{y})^2$
28	-3.286	10.8	29	0.84	0.7056
30	-1.286	1.6538	30	1.84	3.3856
32	0.714	0.51	30	1.84	3.3856
33	1.714	2.94	24	-416	17.3056
33	1.714	2.94	27	-1.16	1.3456
29	-2.286	5.226	29	0.84	0.7056
34	2.714	7.366			
$\sum x = 21$		$\sum (x - \bar{x})^2 = 31.43$	$\sum y = 16$		$\sum (y - \bar{y})^2 = 26.83$

Now $S^2 = \frac{1}{n_1 + n_2 - 2}[(\sum (x - \bar{x})^2 + \sum (y - y)^2]$

$= \frac{1}{11}[31.4358 + 26.8336]$

$= \frac{1}{11}(58.2694)$

= 5.23

∴ S = $\sqrt{5.23}$ = 2.3

i) Null Hypothesis H_0: $\mu_1 = \mu_2$

ii) Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$

iii) Test Statistic : $t_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$= \frac{31.286 - 28.16}{2.3\left(\sqrt{\frac{1}{7} + \frac{1}{6}}\right)} = 2.443$$

∴ t_{cal} = 2.443

iv)Degrees of freedom = 7+6-2 =11

Tabulated value of t for 11 d.f at 5% los is 2.2

Conclusion: Since $|t_{cal}|$ $value$ > t_α value , we reject H_0

Hence we conclude that both horses do not have the same running capacity.

6. Ten soldiers participated in a shooting competition in the first week. After intensive training they participated in the competition in the second week . Their scores before and after training are given below :

Scores before	67	24	57	55	63	54	56	68	33	43
Scores after	70	38	58	58	56	67	68	75	42	38

Do the data indicate that the soldiers have been benefited by the training.

Sol: Given $n_1 = 10$, $n_2 = 10$

We first compute the sample means and standard deviations

$\bar{x}$ = Mean of the first sample = $\frac{1}{10}$ (67 + 24 + 57 +55+63+54+56+68+33+43)

$= \frac{1}{10}(520) = 52$

$\bar{y}$ = Mean of the second sample = $\frac{1}{10}$ (70+38+58+58+56+67+68+75+42+38)

$= \frac{1}{10}(570) = 57$

x	$x - \bar{x}$	$(x - \bar{x})^2$	y	$y - \bar{y}$	$(y - \bar{y})^2$
67	15	225	70	13	169
24	-28	784	38	-19	361
57	5	25	58	1	1
55	3	9	58	1	1
63	11	121	56	-1	1
54	2	4	67	10	100
56	4	16	68	11	121
68	16	256	75	18	324
33	-19	361	42	-15	225
43	-9	81	38	-19	361
$\sum x = 520$		$\sum (x - \bar{x})^2 = 188$	$\sum y = 570$		$\sum (y - \bar{y})^2 = 166$

Now $S^2 = \frac{1}{n_1 + n_2 - 2}[(\sum (x - \bar{x})^2 + \sum (y - y)^2]$

$= \frac{1}{18}[1882 + 1664]$

$= \frac{1}{18}(3546)$

= 197

$\therefore$ S = $\sqrt{197}$ = 14.0357

i) Null Hypothesis H_0: $\mu_1 = \mu_2$

ii) Alternative Hypothesis H_1 : $\mu_1 < \mu_2$ (Left one tailed test)

iii) Test Statistic : $t_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

$$= \frac{52 - 57}{14.0357\left(\sqrt{\frac{1}{10}+\frac{1}{10}}\right)}$$

$= \frac{3546}{18} = -0.796$

$\therefore$ t_{cal} = -0.796

iv)Degrees of freedom = 10+10-2 =18

Tabulated value of t for 18 d.f at 5% los is -1.734

Conclusion: Since $|t_{cal}|\ value < |t_\alpha|$ value , we accept H_0

Hence we conclude that the soldiers are not benefited by the training.

7. The blood pressure of 5 women before and after intake of a certain 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% los?

Sol Given n = 5

i) Null Hypothesis H_0: $\mu_1 = \mu_2$

ii) Alternative Hypothesis H_1 : $\mu_1 < \mu_2$ (Left one tailed test)

iii) Test Statistic $t_{cal} = \frac{\bar{d}}{s/\sqrt{n}}$

where $\bar{d} = \frac{\sum d}{n}$ and $S^2 = \frac{1}{n-1}\sum(d - \bar{d})^2$

B.P before training	B.P after training	$d = y - x$	$d - \bar{d}$	$(d - \bar{d})^2$
110	120	10	8	64
120	118	-2	-4	16
123	125	2	0	0
132	136	4	2	4

125	121	-4	-6	36
		$\sum d = 10$		$\Sigma(d - \bar{d})^2 =$ 120

$\therefore \overline{d} = \frac{10}{5} = 2$ and $S^2 = \frac{120}{4} = 30$

$\therefore$ S = 5.477

$t_{cal} = \frac{\bar{d}}{s/\sqrt{n}} = \frac{2}{5.477/\sqrt{5}} = 0.862$

iv) Degrees of freedom = 5-1= 4

Tabulated value of t for 4 d.f at 1% los is 4.6

Conclusion: Since $|t_{cal}|\ value < |t_\alpha|$ value , we accept H_0

Hence we conclude that there is no significant difference in Blood pressure after intake of a certain drug.

8. Memory capacity of 10 students were tested before and after training . State whether the training was effective or not from the following scores.

Sol : i) Null Hypothesis H_0: $\mu_1 = \mu_2$

ii) Alternative Hypothesis $H_1 : \mu_1 < \mu_2$ (Left one tailed test)

iii) Test Statistic $t_{cal} = \frac{\bar{d}}{s/\sqrt{n}}$

where $\bar{d} = \frac{\Sigma d}{n}$ and $S^2 = \frac{1}{n-1}\Sigma(d - \bar{d})^2$

Before(x)	After(y)	$d = y - x$	d^2
12	15	-3	9
14	16	-2	4
11	10	1	1
8	7	1	1
7	5	2	4
10	12	-2	4
3	10	-7	49
0	2	-2	4
5	3	2	4

6	8	-2	4
		$\sum d = -12$	$\sum d^2 = 84$

$\bar{d} = \frac{-12}{10}$ = -1.2

$S^2 = \frac{84-(-1.2)^2 \times 10}{9} = 7.73$

∴ S = 2.78

$t_{cal} = \frac{\bar{d}}{s/\sqrt{n}} = \frac{-1.2}{2.78/\sqrt{10}}$ = -1.365 and d.f = n-1 = 9

Tabulated value of t for 9 d.f at 5% los is 1.833

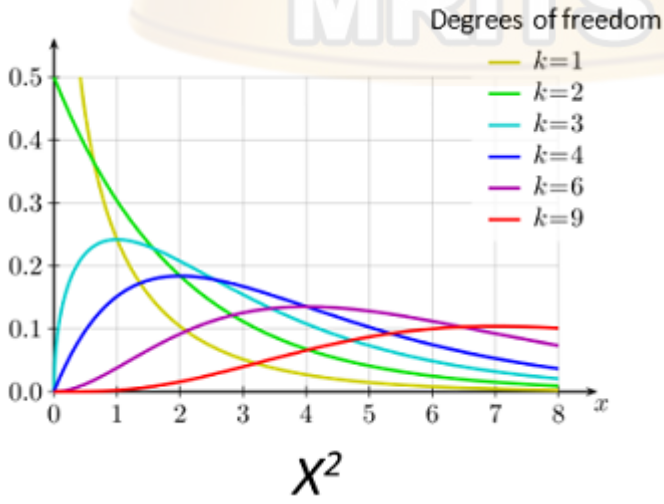
Conclusion: Since $|t_{cal}|\ value < |t_{\alpha}|$ value , we accept H_0

Hence we conclude that there is no significant difference in memory capacity after the training program.

CHI-SQUARE χ^2 TEST

Chi square distribution is a type of cumulative probability distribution . probability distributions provide the probability of every possible value that may occur . Distributions that are cumulative give the probability of a random variable being less than or equal to a particular value. Since the sum of the probabilities of every possible value must equal one , the total area under the curve is equal to one . Chi square distributions vary depending on the degrees of freedom. The degrees of freedom is found by subtracting one from the number of categories in the data .

.



X^2

Applications of Chi – Square Distribution:

Chi – Square test as a test of goodness of fit :

χ^2– test enables us to ascertain how well the theoretical distributions such as binomial, Poisson, normal etc, fit the distributions obtained from sample data. If the calculated value of $\chi^2$is less than the table value at a specified level of generally 5% significance, the fit is considered to be good.

If the calculated value of χ^2 is greater than the table value, the fit is considered to be poor.

i) **Null hypothesis**: H_0 : There is no difference in given values and calculated values

ii)**Alternative hypothesis**: H_1 : There is some difference in given values and calculated values

iii) **Test Statistic** $\chi^2_{cal} = \sum \frac{(O-E)^2}{E}$

iv)At specified level of significance for n-1 d.f if the given problem is binomial distribution

At specified level of significance for n-2 d.f if the given problem is Poisson distribution

v)**Conclusion** :If ψ^2_{cal} value < ψ^2_{tab} value , then we accept H_0 , Otherwise reject H_0.

2. **Chi – Square test for independence of attributes :**

Definition : An attribute means a quality or characteristic
Eg: Drinking, Smoking, blindness, Honesty, beauty etc.,

An attribute may be marked by its presence or absence in a number of a given population.

Let us consider two attributes A and B.

A is divided into two classes and B is divided into two classes. The various cell frequencies can be expressed in the following table known as 2x2 contingency table.

A	**a**	**b**	**a+b**
	c	**d**	**c+d**
	a+c	**b+d**	**N=a+b+c+d**

The expected frequencies are given by

$$E(a) = \frac{(a+c)(a+b)}{N}$$

$$E(b) = \frac{(b+c)(a+b)}{N}$$

$$E(c) = \frac{(a+c)(c+d)}{N}$$

$$E(d) = \frac{(b+d)(c+d)}{N}$$

$$\chi^2_{cal} = \sum \frac{(O-E)^2}{E}$$

χ^2_{cal} value to be compared with χ^2_{tab} value at 1% (5.1 or10%) level of significance for (r-1) (c-1) d.f where r- number of rows

c-number of columns.

Note: In χ^2 distribution for independence of attributes, we test if two attributes A and B are independent or not.

i)**Null Hypothesis**: H_0 : The two attributes are independent

ii) **Alternative hypothesis**: H_1 : The two attributes are not independent

iii) **Test Statistic** $\chi^2_{cal} = \sum \frac{(O-E)^2}{E}$

where E = $\frac{\text{Row total x Column total}}{\text{Grand total}}$

iv)At specified level of significance for (m-1) (n-1) d.f where m- no. of rows and n- no. of columns

v)**Conclusion** : If χ^2_{cal} value < χ^2_{tab} value , then we accept H_0 , Otherwise reject H_0.

Problems :

1. Fit a Poisson distribution to the following data and test for its goodness of fit at 5% los

x	0	1	2	3	4
f	419	352	154	56	19

Sol:

X	f	f.x

0	419	0
1	352	352
2	154	308
3	56	168
4	19	76
	N=1000	$\sum fx = 904$

Mean $\lambda = \frac{\Sigma fx}{N} = \frac{904}{1000} = 0.904$

Theoretical distribution is given by

$= N \times p(x) = 1000 \times \frac{e^{-\lambda}\lambda^x}{x!}$

Hence the theoretical frequencies are given by

x	0	1	2	3	4	Total
$f = 1000 \times \frac{e^{-\lambda}\lambda^x}{x!}$	406.2	366	165.4	49.8	12.6	1000

Since Given frequencies total is equal to Calculated frequencies total.

To test for goodness of fit:

i) H_0 : There is no difference in given values and calculated values

ii) H_1 : There is some difference in given values and calculated values

iii) $\chi^2_{cal} = \sum \frac{(O-E)^2}{E}$

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
419	406.2	$(419-406.2)^2$	$\frac{(419-406.2)^2}{406.2}$
352	366	$(352-366)^2$	$\frac{(352-366)^2}{366}$
154	165.4	$(154-165.4)^2$	$\frac{(154-165.4)^2}{165.4}$

56	49.8	$(56-49.8)^2$	$\frac{(56-49.8)^2}{49.8}$
19	12.6	$(19-12.6)^2$	$\frac{(19-12.6)^2}{12.6}$

$\sum\frac{(O-E)^2}{E}$ = 5.748

Degrees of freedom = 5-2 = 3

χ^2_{tab} at 5% LOS = 7.82

Since χ^2_{cal} value < χ^2_{tab} ,we accept H_0 .

3. A die is thrown 264 times with following results. Show that the die is biased [Given $\chi^2_{0.05}$ = 11.07 for 5 d.f]

No. appeared on the die	1	2	3	4	5	6
Frequency	40	32	28	58	54	52

Sol: i) H_0 : The die is unbiased

ii) H_1 : The die is not unbiased

iii)$\chi^2_{cal} = \sum\frac{(O-E)^2}{E}$

The expected frequency of each of the number 1,2,3,4,5,6 is $\frac{264}{6} = 44$

Calculation of χ^2:

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
40	44	16	0.3636
32	44	144	3.2727
28	44	256	5.8181
58	44	196	4.4545

54	44	100	2.2727
52	44	64	1.4545

$\sum \frac{(O-E)^2}{E}$ = 17.6362

χ^2_{cal} = 17.6362

The number of degrees of freedom = n-1 = 5

$\chi^2_{0.05}$ = 11.07 for 5 d.f

Since χ^2_{cal} value > χ^2_{tab} value , we reject H_0

Hence the die is biased

4. On the basis of information given below about the treatment of 200 patients suffering from disease , state whether the new treatment is comparatively
Superior to the conventional treatment.

Treatment	Favorable	Not Favorable	Total
New	60	30	90
Conventional	40	70	110

Sol: i) H_0 : The two attributes are independent

ii) H_1 : The two attributes are not independent

iii) $\chi^2_{cal} = \sum \frac{(O-E)^2}{E}$

where E = $\frac{\text{Row total x Column total}}{\text{Grand total}}$

$\frac{90 \times 100}{200} = 45$	$\frac{90 \times 100}{200} = 45$	90
$\frac{100 \times 110}{200} = 55$	$\frac{100 \times 110}{200} = 55$	11
100	100	200

Calculation of χ^2:

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
60	45	225	5

30	45	225	5
40	55	225	4.09
70	55	225	4.09

$\sum \frac{(O-E)^2}{E} = 18.18$

$\chi^2_{cal} = 18.18$

χ^2_{tab} for 1 d.f . at 5% los is 3.841

since χ^2_{cal} value > χ^2_{tab} value , we reject H_0

Hence we conclude that new and conventional treatment are not independent.

SNEDECOR'S F- TESTOF SIGNIFICANCE:

The **F-Distribution** is also called as **Variance Ratio Distribution** as it usually defines the ratio of the variances of the two normally distributed populations. The F-distribution got its name after the name of **R.A. Fisher,** who studied this test for the first time in 1924.

Symbolically, the quantity is distributed as F-distribution with and degrees of freedom $\vartheta_1 = n_1 - 1$and $\vartheta_2 = n_2 - 1$ is represented as:

$$F_{cal} = \frac{Greater\ Variance}{Smaller\ Varinace}$$

$$F_{cal} = \frac{S_1^2}{S_2^2} \text{ Or } \frac{S_2^2}{S_1^2}$$

Where,

S_1^2 is the unbiased estimator of σ_1^2 and is calculated as: $S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{1}{n_1 - 1} \sum (x_1 - \overline{x_1})^2$

S_2^2 is the unbiased estimator of σ_2^2 and is calculated as: $S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{1}{n_2 - 1} \sum (x_2 - \overline{x_2})^2$

To test the hypothesis that the two population variances σ_1^2 and σ_2^2 are equal

i) $H_0 : \sigma_1^2 = \sigma_2^2$

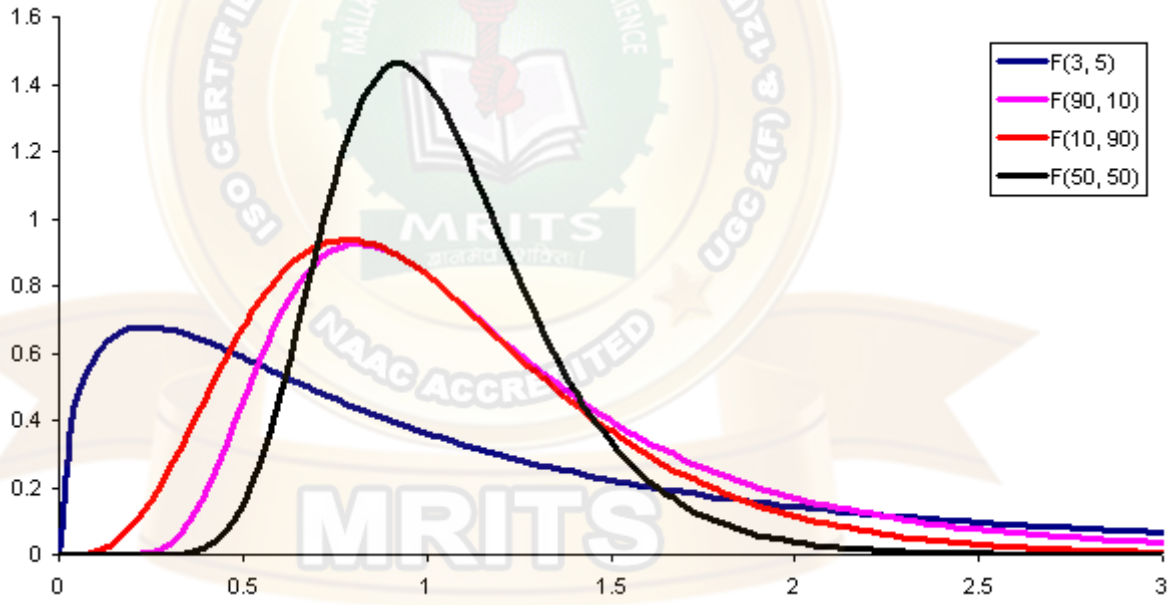
ii) $H_1 : {\sigma_1}^2 \neq {\sigma_2}^2$

iii) $F_{cal} = \frac{\textbf{Greater Variance}}{\textbf{Smaller Varinace}}$

iv)At specified level of significance (1% or 5 %) for $(\vartheta_1, \vartheta_2)$ d.f

v) If $\mathbf{F_{cal}}$ value < $\mathbf{F_{tab}}$ value , then we accept H_0 , Otherwise reject H_0.

$F_{cal}(\vartheta_1, \vartheta_2)$ is the value of F with ϑ_1 and ϑ_2 degrees of freedom such that the area under the F – distribution to the right of F_α is α.



Problems:

1. In one sample of 8 observations from a normal population, the sum of the squares of deviations of the sample values from the sample mean is 84.4 and in another sample of 10 observations it was 102.6. Test at 5% level whether the populations have the same varience.

Sol: Let ${\sigma_1}^2$ and ${\sigma_2}^2$ be the variances of the two normal populations from which the samples are drawn.

Let the Null Hypothesis be $H_0: {\sigma_1}^2 = {\sigma_2}^2$

Then the Alternative Hypothesis is $H_1: {\sigma_1}^2 \neq {\sigma_2}^2$

Here $n_1 = 8, n_2 = 10$

Also $\sum(x_i - \bar{x})^2 = 84.4, \sum(y_i - \bar{y})^2 = 102.6$

If ${S_1}^2 and {S_2}^2$ be the estimates of ${\sigma_1}^2$ and ${\sigma_2}^2$ then

$${S_1}^2 = \frac{1}{n_1 - 1}\sum(x_i - \bar{x})^2 = \frac{84.4}{7} = 12.057$$

and

$${S_2}^2 = \frac{1}{n_2 - 1}\sum(y_i - \bar{y})^2 = \frac{102.6}{9} = 11.4$$

Let H_0 be true. Since ${S_1}^2 > {S_2}^2$, the test statistic is

$$F = \frac{{S_1}^2}{{S_2}^2} = \frac{12.057}{11.4} = 1.057$$

i.e., calculated F = 1.057.

Degrees of freedom are given by $v_1 = n_1 - 1 = 8 - 1 = 7$ and $v_2 = n_2 - 1 = 10 - 1 = 9$

Tabulated value of F at 5% level for $(7,9)$ degrees of freedom is 3.29

i.e.,$F_{0.05}(7,9) = 3.29$

Since calculated $F <$ tabulated F, we accept the Null Hypothesis H_0 and conclude that the populations have the same variance.

2.The time taken by workers in performing a job by method I and method II is given below:

Method I	20	16	26	27	23	22	-
Method II	27	33	42	35	32	34	38

Do the data show that the variances of time distribution from population from which these samples are drawn do not differ significantly?

Sol: Let the Null Hypothesis be $H_0: {\sigma_1}^2 = {\sigma_2}^2$ where ${\sigma_1}^2$ and ${\sigma_2}^2$ are the variances of the two populations from with the samples are drawn.

The Alternative Hypothesis is $H_1: {\sigma_1}^2 \neq {\sigma_2}^2$.

Calculation of sample variances.

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
20	-2.3	5.29	27	-7.4	54.76
16	-6.3	39.69	33	-1.4	1.96
26	3.7	13.69	42	7.6	57.76
27	4.7	22.09	35	0.6	0.36
23	0.7	0.49	32	-2.4	5.76
22	-0.3	0.09	34	-0.4	0.16
			38	3.6	12.96
134		81.34	241		133.72

Given $n_1 = 6, n_2 = 7$

$$\therefore \bar{x} = \frac{\sum x}{n_1} = \frac{134}{6} = 22.3, \bar{y} = \frac{\sum y}{n_2} = \frac{241}{67} = 34.4$$

And $\sum(x_i - \bar{x})^2 = 81.34, \sum(y_i - \bar{y})^2 = 133.72$

If ${S_1}^2 and {S_2}^2$ be the estimates of ${\sigma_1}^2$ and ${\sigma_2}^2$, then

$${S_1}^2 = \frac{1}{n_1 - 1}\sum(x_i - \bar{x})^2 = \frac{81.34}{5} = 16.26$$

and

$${S_2}^2 = \frac{1}{n_2 - 1}\sum(y_i - \bar{y})^2 = \frac{133.72}{6} = 22.29$$

Let H_0 be true

Since ${S_2}^2 > {S_1}^2$, the statistic is

$$F = \frac{{S_2}^2}{{S_1}^2} = \frac{22.29}{16.268} = 1.3699 = 1.37$$

$F_{0.05}(5,6)$ d.f $= 4.39$

Since calculated F < tabulated F , we accept the null hypothesis H_0 at 5% los i.e., there is no significant difference between the variances of the distribution by the workers.

Unit - VIII

Stochastic Processes

Stochastic Processes :- A Stochastic processes is a set of random variables $\{X(t); t \in T\}$ depending on some real parameters like time "t", where $t \geq 0$. the stochastic Processes is also known as random processes. It defines the random variables whose outcomes continuously changes with time.

Classification of stochastic processes:-

(i) **continuous stochastic processes:-**

If the process is defined for all instants of time (i.e $t \geq 0$), then it is called continuous stochastic process and represented by $\{X(t), t \geq 0\}$ or If both x and t are continuous the stochastic process is called as continuous stochastic process.

(ii) **Discrete stochastic process:-**

If the process is defined for discrete time instants (i.e, $n = 1, 2 \cdots$) then it is called discrete stochastic process. It is represented by $\{x_n, n = 1, 2 \cdots\}$

Stationary Random process :-

A random process $x(t)$ is said to be stationary or strict-sense stationary (sss) if all the statistical properties of the random process are not affected by a shift in the time. i.e fo

Markov process (Markov model) is the value of the process depends only upon the most recent previous values and is independent of all values in the more distant past.

Markov process is an effective and powerful tool for prediction of share market since this process assumes prob values, which yields better predicted values.

Random process is a collection of r.v. $\{x(s,t)\}$ that are functions of a r.v time t.,

where $s \in$ sample space, $t \in T$.

State Space: is the set of possible values of any individual member of the random process.

If the parameter set 'T' is discrete, the random process will denoted by $\{X_n\}$.

If the parameter set 'T' is continuous, the process will be denoted by $\{X(t)\}$.

note: In Markov process., the future value does not depend on the past values, but only on the present value

A random process $\{X(t)\}$ is said to be

Markovian if

$$P\left[X(t_{n+1}) \le x_{n+1} \mid X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \ldots X(t_0) = x_0\right] = P\left[X(t_{n+1}) \le x_{n+1} \mid X(t_n) = x_n\right]$$

where $t_0 \le t_1 \le t_2 \le \ldots t_n \le t_{n+1}$.

Here $x_0, x_1, x_2 \ldots x_n, x_{n+1}$ are called the states of the process.

If the random process at time t_n is in the state x_n, the future state of the random process x_{n+1} at t_{n+1} depends only on the present state x_n and not on the past states.

eg:). The prob. of raining today depends only on previous weather conditions existed for the last two days and not on past weather conditions is a markovian process.

(2) If Airplane departed now is of certain airline, then there is less prob. of having next airplane from same airline. That means if we know the airplane departed then we can predict the prob. of airplane from certain airline

Markov chain:- Let $\{X(t)\}$ be a Markov process that possess markov property and which takes only discrete values whether it is discrete (or) conti., then $\{X(t)\}$ is called as Markov chain

Mathematically, we define Markov chain as follows.

If $P\{X_n = a_n / X_{n-1} = a_{n-1}, X_{n-2} = a_{n-2}, \ldots X_0 = a_0\}$

$= P\{X_n = a_n / X_{n-1} = a_{n-1}\}$

for all n,

then the process $\{X_n\}$, $n = 0, 1, 2, \ldots$ is called as Markov chain.

Here $a_0, a_1, a_2, \ldots a_n$ are called states.

One-step transition prob:- The condition prob. $P\{X_n = a_j / X_{n-1} = a_i\}$ is called the 1-step tran. prob. from state a_i to state a_j at the nth step (trial) and it denoted by $P_{ij}(n-1, n)$.

$$P_{ij}(n-1, n) = P\{X_n = a_j / X_{n-1} = a_i\}.$$

Homogenous Markov chain:-

If $P_{ij}(n-1, n) = P_{ij}(m-1, m)$ then the Markov Chain is called the homo. M.C. (or) the Chain is said to have stationary transition probabilities.

Transition prob. Matrix (TPM) :- when the Markov chain is homogeneous, the 1-step transition prob. denoted by $P = \{P_{ij}\}$ is called 1-step TPM.

n-step transition prob:- The conditional prob. that the process is in state a_j at step n given that it was in state a_i at step 0,

ie, $P\{X_n = a_j / X_0 = a_i\}$ is called the n-step transition prob. and is defined as

$$P = [P_{ij}^{n}] = P\{X_n = a_j / X_0 = a_i\}$$

Chapman-Kolmogorov:-

If P is the transition prob. matrix of a homo. Markov chain, then the n-step tran. prob. Matrix $P^{(n)}$ is equal to P^n.

ie, $[P_{ij}^{(n)}] = [P_{ij}]^n$.

Regular Matrix:- A stochastic matrix P is said to be a regular matrix, if all the entries of P^m (for some +ve int m) are positive.

A homogenous Markov chain is said to be regular if its transition prob. matrix is regular.

Markov chains:— Consider a system which can be in any one of a finite no. of states $E_1, E_2, \ldots E_n$. We also assume that the prob. of the system being in a given state at the next trial depends only on its present state and not upon the states it may have been in earlier times.

If at any time, the system is in a state E_i, the prob. of it being in state E_j at next trial is P_{ij}.

This process is known as Markov chain and this property of the process is called Markov property.

P_{ij} is called the prob. of a transition from E_i to E_j.

Stochastic matrix:— The transition probabilities P_{ij} will be arranged in a matrix of transition probabilities

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} \cdots \\ P_{21} & P_{22} & P_{23} \cdots \\ \cdots & & \end{bmatrix}$$

where the first subscript stands for row,
the second " " " column.

Clearly P is a square matrix with non-negative elements and sum of the elements in each row is equal to 1.

Such a matrix (finite/infinite) is called a stochastic matrix.

Th: If P, Q are stochastic matrices then product PQ is also a stochastic matrix.

Thus P^n is a stochastic matrix for all the int. values of n.

→ A stochastic matrix P is said to be regular if all the entries of some power P^n are positive.

→ A " " " " not regular if 1 occur in the principle main diagonal.

eg ① which of the following matrices are stochastic

(i) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}_{2\times 3}$ ii) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ iii) $\begin{bmatrix} 0 & 1 \\ 1/3 & 1/4 \end{bmatrix}$ iv) $\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$

v) $\begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}$

Sol: i) is not a square matrix

∴ It is not stochastic

ii) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ The matrix is a square matrix with non-negative entries and sum of the elements in each row is equal to 1.

∴ The matrix is stochastic.

iii) $\begin{bmatrix} 0 & 1 \\ 1/3 & 1/4 \end{bmatrix}$ is a square matrix but sum in each row is not equal to 1.

∴ It is not stochastic

iv) $\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$ is a sq. matrix with sum of the elements in each row is 1

∴ stochastic

v) $\begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}$: It contains negative elements

⇒ It is not stochastic.

(7) Which of the stochastic matrices are regular.

(i) $\begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}$

It is not regular since 1 lies on the main diagonal.

(ii) $A = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$

$A^2 = AA = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 3/8 & 3/8 & 1/4 \end{bmatrix}$, $A^3 = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 7/16 & 7/16 & 1/8 \end{bmatrix}$

Since the entries a_{13}, a_{23} are zero.

A is not regular.

(iii) $B = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$

$B^2 = B \cdot B = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix}$, $B^3 = B \cdot B^2 = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{bmatrix}$

$B^4 = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/2 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}$, $B^5 = \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/8 & 1/2 & 3/8 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$

Since all the entries of some power of B are positive, B is regular stochastic matrix.

Examples of markov chains:-

Eg:1:- All the transition probabilities can be collected into a matrix, $P = P_{ij}$ called transition probability matrix.

$$P = \begin{bmatrix} P_{00} & P_{01} & P_{02} \cdots \\ P_{10} & P_{11} & P_{12} \cdots \\ \vdots & \vdots & \vdots \\ P_{i0} & P_{i1} & P_{i2} \cdots \\ \vdots & \vdots & \vdots \end{bmatrix}$$

Eg:2:-

A communication channel:- Starting at time zero, a binary symbol (0 or 1) is transmitted every second across a simple communication channel to a receiver. The probability of successful transmission (i.e $0 \to 0$ or $1 \to 1$) is $(1-p)$ while the probability of error (i.e $0 \to 1$ or $1 \to 0$) is p where, $0 \le p \le 1$.

We can model this as a discrete time markov chain with state $S = \{0, 1\}$, the transition matrix is

$$P = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}.$$

Eg:-3:-

Random walk model:-

A random walk is an example of a markov chain where the state space is genuinely infinite. To make it easier to write down P, lets impose the 'initial condition' $P_{01} = 1$. For $i > 0$ the only non-zero transition probabilities are

$$P_{i,i+1} = p, \quad P_{i,i-1} = 1-p$$

The transition matrix is;

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 1-P & 0 & P & 0 & \cdots \\ 0 & 1-P & 0 & P & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

Eg:- 4):-

A Gambling model:— consider a gambler wins Rs.1. with probability p or loses with probability (1-p) as long as he plays.

We can model this as a markov chain with state space $S = \{0, 1, \cdots N\}$ and transition probabilities $P_{00} = P_{NN} = 1$ and $P_{i,i+1} = P$, $P_{i,i-1} = 1-P$.

Because of similarities with the Example-3, this is called a finite state random walk.

When N=4, the transition matrix takes the following form.

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1-P & 0 & P & 0 & 0 \\ 0 & 1-P & 0 & P & 0 \\ 0 & 0 & 1-P & 0 & P \\ 0 & 0 & 0 & 1-P & 1 \end{bmatrix}.$$

Chapman-Kolmogrov equation: — For a markov chain $\{x_n; n \geq 0\}$ defined on the state space $S = \{0, 1, 2 \ldots\}$ then $P_{ij}^{(m+n)} = \sum_{k \in S} P_{ik}^{(m)} \cdot P_{kj}^{(n)}$ for m, n positive integer.

Classification of states: —

Recurrent state :— A state 'i' is called recurrent if $F_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)} = 1$. (or) $\sum_{n=1}^{\infty} P_{ij}^{(n)} = 1$

Transient state: — A state 'i' is called transient if $F_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)} < 1$. (or) $\sum_{n=1}^{\infty} P_{ij}^{(n)} < 1$

Irreducible :— If every state can be reached from any state then the chain is called irreducible. Then the transition matrix is irreducible.

Absorbing state: — A state 'i' is said to be an absorbing state if and only if $P_{ii} = 1$.

A markov chain is absorbing if it has at least on absorbing state and it is possible to go from every non-absorbing state to atleast one absorbing state in one or more steps.

Periodic :— A state is said to be periodic with period 't' if the return to state is possible only in finite steps.

The state 'i' is said to be aperiodic (or non-periodic) of no such 't' exists.

Ergodic :— A positive recurrent and aperiodic state is called Ergodic. A Markov chain all of whose states are ergodic is set to be a "ergodic chain".

1) The tansition probability matrix is given by

$$P = \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.2 & 0.2 & 0.6 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} \quad \text{and} \quad P_0 = [0.4, 0.4, 0.2]$$

a) Find the distribution after Three transitions.

b) Find the limiting probabilities.

Sol:- (a) The distribution after three transitions is

$P_0 P^3$

$$\therefore P^2 = P \cdot P = \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.2 & 0.2 & 0.6 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.2 & 0.2 & 0.6 \\ 0.7 & 0.2 & 0.1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.44 & 0.22 & 0.34 \\ 0.48 & 0.24 & 0.28 \\ 0.18 & 0.34 & 0.48 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 0.44 & 0.22 & 0.34 \\ 0.48 & 0.24 & 0.28 \\ 0.18 & 0.34 & 0.48 \end{bmatrix} \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.2 & 0.2 & 0.6 \\ 0.7 & 0.2 & 0.1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.326 & 0.288 & 0.386 \\ 0.292 & 0.296 & 0.412 \\ 0.422 & 0.236 & 0.342 \end{bmatrix}$$

$$\therefore P_0P^3 = [0.4 \quad 0.4 \quad 0.2] \begin{bmatrix} 0.326 & 0.288 & 0.386 \\ 0.292 & 0.296 & 0.412 \\ 0.422 & 0.236 & 0.342 \end{bmatrix}$$

$$= [0.3316 \quad 0.2808 \quad 0.3876]$$

(b) we have to find limiting probabilities

using $XP = X$

where $X = (x \ y \ z)$

$$\therefore XP = X$$

$$(x \ y \ z) \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.2 & 0.2 & 0.6 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} = (x \ y \ z)$$

$$\Rightarrow 0.1x + 0.2y + 0.7z = x$$

$$0.4x + 0.2y + 0.2z = y$$

$$0.5x + 0.6y + 0.1z = z$$

& $x + y + z = 1$

solving above eqn we get

$x = 0.3529, \quad y = 0.2706, \quad z = 0.3765.$

3) The transition probability matrix of a markov chain $\{x_n\}$; $n=1,2,3\cdots$ having three states 1, 2 & 3 is

$$P = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$$ and the initial distribution is $p^{(0)} = (0.7 \quad 0.2 \quad 0.1)$

Find (i) $P\{x_2=3\}$

(ii) $P\{x_3=2, x_2=3, x_1=3, x_0=2\}$

Sol:- $$p^{(2)} = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$$

$$= \begin{bmatrix} 0.43 & 0.31 & 0.26 \\ 0.24 & 0.42 & 0.34 \\ 0.36 & 0.35 & 0.29 \end{bmatrix}$$

(i) $$P\{x_2=3\} = \sum_{i=1}^{3} P\{x_2=3/x_0=i\} \cdot P(x_0=i)$$

$$= P\{x_2=3/x_0=1\}\, P(x_0=1) + P\{x_2=3/x_0=2\}\, P(x_0=2) + P\{x_2=3/x_0=3\}\, P(x_0=3)$$

$$= p_{13}^{(2)}\, p_1^{(0)} + p_{23}^{(2)}\, p_2^{(0)} + p_{33}^{(2)}\, p_3^{(0)}$$

$$= 0.26 \times 0.7 + 0.34 \times 0.2 + 0.29 \times 0.1$$

$$= 0.279$$

=

(iii) $P\{X_3=2, X_2=3, X_1=3, X_0=2\}$

Now $P\{X_1=3/X_0=2\} = P_{23} = 0.2$

$P\{X_1=3, X_0=2\} = P\{X_1=3/X_0=2\} \cdot P\{X_0=2\}$

$= 0.2 \times 0.2 = 0.04$

$P\{X_2=3, X_1=3, X_0=2\} = P\{X_2=3/X_1=3, X_0=2\} \cdot P\{X_1=3, X_0=2\}$

$= P\{X_2=3/X_1=3\} \times P\{X_1=3, X_0=2\}$ [∵ markov property]

$= 0.3 \times 0.04 = 0.012$

$P[X_3=2, X_2=3, X_1=3, X_0=2\} = P\{X_3=2/X_2=3, X_1=3, X_0=2\} \cdot P\{X_2=3, X_1=3, X_0=2\}$

$= P\{X_3=2/X_2=3\} \cdot P\{X_2=3, X_1=3, X_0=2\}$

$= 0.4 \times 0.012$

$= 0.0048$

⇐

4) The three state markov chain is given by the the transition prob^y matrix

$$P = \begin{matrix} \\ 0 \\ 1 \\ 2 \end{matrix} \begin{matrix} \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} \end{matrix}$$

P.T the chain is irreducible.

Sol:- $P^2 = P \cdot P = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$

$$= \begin{bmatrix} 3/6 & 1/6 & 1/3 \\ 1/4 & 7/12 & 1/6 \\ 1/4 & 1/3 & 5/12 \end{bmatrix}$$

$P_{00}^{(2)} > 0$, $P_{01}^{(1)} > 0$, $P_{02}^{(2)} > 0$

$P_{10}^{(1)} > 0$, $P_{11}^{(2)} > 0$, $P_{12}^{(1)} > 0$

$P_{20}^{(2)} > 0$, $P_{21}^{(1)} > 0$, $P_{22}^{(2)} > 0$

$\therefore$ The chain is irreducible & all the states recurrent.

(or) In this markov chain, all the states communicate with each other.

Suppose we consider the states as 0, 1, 2

It is possible to go from state

0 to state 1 with prob^y $1/2$

Again it is possible to go from state

1 to state 2 with prob^y $1/3$

Thus it is possible to go from state

0 to state 2.

So, chain is irreducible & all the states are recurrent.

=

4 Define a Regular transition matrix.

A Transition matrix P is called Regular, if for some integer n, all entries of P^n are strictly positive.

5 Is $P = \begin{bmatrix} 1 & 0 \\ 0.5 & 0.5 \end{bmatrix}$ a stochastic matrix?

Sol This is a square matrix.

All the entries are ≥ 0.

The sum of the entries in each row is 1.

$\therefore P$ is stochastic

6 Define Recurrent state of Markov chain

Sol A state i is said to be recurrent state, if the system will return to it after leaving some time in the future.

7. If $\begin{bmatrix} 0.5 & x & 0 \\ 0.2 & 0 & x+y \\ z & 0.4 & 0.1 \end{bmatrix}$ is a transmission prob. matrix, then find the values of x, y, z

Sol $0.5 + x = 1 \Rightarrow x = 0.5$

$0.2 + x + y = 1 \Rightarrow y = 0.3$

$z + 0.4 + 0.1 = 1 \Rightarrow z = 0.5$

1 If $\begin{bmatrix} 0.5 & x \\ y & 0.124 \end{bmatrix}$ is Transition probability matrix, then find the values of x and y.

Sol: Since sum of row probabilities is equal to 1.

$0.5 + x = 1 \Rightarrow x = 0.5$

$y + 0.124 = 1 \Rightarrow y = 0.876$.

2 Define continuous random process.

The values assumed by the r.v are called States.

The set of all possible values of an individuals r.v. X_n of a stochastic process $\{X_n, n \geq 1\}$ is known as its State space. It is denoted by I.

The state space is said to be discrete if it contains a finite or countable infinity of points, otherwise it is called continuous random process.

3. what do you mean by Stochastic matrix. Give an example.

A Stochastic matrix is a random matrix with non-negative elements and unit row sums.

eg: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is a Stochastic since each row sum is 1.

A training process is considered as a two state marchov chain. If it rains it is considered as 0, if not 1. The transit in prob. matrix of the marcov chain is $\begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$. Find the prob. that it will rain after 3 days, assuming that the initial probabilities are 0.4 & 0.6.

Sol:

$$P_0 = (0.4 \quad 0.6), \quad P = \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}$$

$$P^{(1)} = P_0 P = (0.4 \quad 0.6) \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix} = (0.36 \quad 0.64)$$

$$P^{(2)} = P^{(1)} P = (0.36 \quad 0.64) \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix} = (0.344 \quad 0.656)$$

$$P^{(3)} = P^{(2)} P = (0.344 \quad 0.656) \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix} = (0.3376 \quad 0.6624)$$

The prob. that there will be rain after 3 days is 0.3376.

In a presidential election at the end of June 40% of the voters registered for liberal, 45% for conservative and 15% for Independent. Over one month, of the people those who registered for liberal 80% were retained, 15% changed to conservative and 5% to Independent.
Of the people those who registered for conservative 70% were retained, 20% changed to liberal and 10% to independent.

of the people those who registered for independent 60% were retained, 20% changed to liberal and 20% to conservative

a) write the transition prob matrix

b) Find the % of the voters in each category at the end july

c) At the end August.

Sol:

$$p_0 = (0.4 \quad 0.45 \quad 0.15)$$

(a) The transition prob matrix $= \begin{pmatrix} 0.8 & 0.15 & 0.05 \\ 0.2 & 0.7 & 0.1 \\ 0.2 & 0.2 & 0.6 \end{pmatrix}$

(b) $p_0 P = (0.4 \quad 0.45 \quad 0.15) \begin{pmatrix} 0.8 & 0.15 & 0.05 \\ 0.2 & 0.7 & 0.1 \\ 0.2 & 0.2 & 0.6 \end{pmatrix}$

$$p^{(1)} = (0.449 \quad 0.405 \quad 0.155)$$

The voters percentage in each category at the end of July

$= 44.9, 40.5, 15.5$ %

(c) $p^{(2)} = p_0 p^{(1)}$

$$= (0.449 \quad 0.405 \quad 0.155) \begin{pmatrix} 0.8 & 0.15 & 0.05 \\ 0.2 & 0.7 & 0.1 \\ 0.2 & 0.2 & 0.6 \end{pmatrix}$$

$$= (0.464 \quad 0.3805, 0.1555)$$

The voters % in each category at the end of Aug $= 46.4, 38.05, 15.55$

If the transition prob. matrix is $\begin{bmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$

and the initial probabilities are $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$

then find:

(a) the probabilities after 3 periods

(b) Equilibrium vector.

Sol:

$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{bmatrix} \end{array}$$

$$p_1^{(0)} = \frac{1}{3}, \quad p_2^{(0)} = \frac{1}{3}, \quad p_3^{(0)} = \frac{1}{3}$$

$$p_0 = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

(a) $p^2 = p \cdot p =$

$p^3 = p^2 \cdot p =$

The probabilities after 3 periods

$= p_0 \cdot p^3 = \left(\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3}\right) \begin{pmatrix} 0.4062 & 0.2031 & 0.3906 \\ 0.4062 & 0.1875 & 0.4062 \\ 0.3906 & 0.2031 & 0.4062 \end{pmatrix}$

$= (0.401 \quad 0.1979 \quad 0.401)$

(b) we find Equilibrium vector V such that

$VP = V$

$VP - V = 0 \Rightarrow V[P - I] = 0 \quad \text{---(1)}$

where $V = (v_1 \; v_2 \; v_3)$

$$P-I = \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} -0.5 & 0.25 & 0.25 \\ 0.5 & -1 & 0.5 \\ 0.25 & 0.25 & -0.5 \end{pmatrix}$$

(4) $\Rightarrow$ $v(P-I) = 0$

$\Rightarrow$ $\begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} (P-I) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

$\Rightarrow$ Solving system of equations

$$-0.5 v_1 + 0.5 v_2 + 0.25 v_3 = 0 \quad \text{---(2)}$$

$$0.25 v_1 - v_2 + 0.25 v_3 = 0 \quad \text{---(3)}.$$

$$0.25 v_1 + 0.5 v_2 - 0.5 v_3 = 0$$

and $v_1 + v_2 + v_3 = 1 \Rightarrow v_1 = 1 - v_2 - v_3$

Solving (2) & (3) by sub. v_1 value

$$v_2 = 0.2$$

$$v_3 = 0.4$$

$\therefore v_1 = 0.4$.

If the transition prob. matrix of market shares of 3 brands A, B and C is $\begin{bmatrix} 0.2 & 0.4 & 0.4 \\ 0.7 & 0.2 & 0.1 \\ 0.2 & 0.3 & 0.4 \end{bmatrix}$ and the initial market shares are 30%, 30%, and 40%. Find

(a) The market shares in 2 and 3rd periods

(b) The limiting probabilities.

Sol:

$P_0 = (0.3 \quad 0.3 \quad 0.4)$

$P^2 = P \cdot P$

$P^3 = P^2 \cdot P$

(a) The market shares in 2nd period is

$= P_0 P^2 = (0.3 \quad 0.3 \quad 0.4) \begin{pmatrix} 0.44 & 0.28 & 0.28 \\ 0.21 & 0.35 & 0.34 \\ 0.39 & 0.3 & 0.31 \end{pmatrix}$

$= (0.321 \quad 0.309 \quad 0.31)$

The market shares in 3rd period is

$= P_0 P^2$

$= (0.3 \quad 0.3 \quad 0.4) \begin{pmatrix} 0.368 & 0.316 & 0.316 \\ 0.409 & 0.296 & 0.295 \\ 0.321 & 0.309 & 0.31 \end{pmatrix}$

$= (0.3655 \quad 0.3072 \quad 0.3073)$

(b) The limiting probabilities

Assume that the transition matrix

is $P = \begin{pmatrix} 0.2 & 0.4 & 0.4 \\ 0.7 & 0.2 & 0.1 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}$

Recall that the n-step transition probabilities are given by powers of P.

So let's look at some powers

$$P^4 = \begin{pmatrix} 0.3896 & 0.3052 & 0.3052 \\ 0.3775 & 0.3113 & 0.3112 \\ 0.3854 & 0.3072 & 0.3073 \end{pmatrix}$$

...

$$P^8 = \begin{pmatrix} 0.3846 & 0.3076 & 0.3076 \\ 0.3845 & 0.3077 & 0.3077 \\ 0.3846 & 0.3076 & 0.3076 \end{pmatrix}$$

$$P^{12} = \begin{pmatrix} 0.3844 & 0.3075 & 0.3075 \\ 0.3845 & 0.3076 & 0.3076 \\ 0.3844 & 0.3075 & 0.3075 \end{pmatrix}$$

and subsequent powers are the same to this problem.

If the transition prob. matrix of a Markov chain is $\begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$, find the steady state distn.

Sol: we find equilibrium vector V such that

$$VP = V$$

$$VP - V = 0 \Rightarrow V(P - I) = 0$$

where $V = (v_1\ v_2)$

$$P - I = \begin{pmatrix} 0 & 1 \\ 0.5 & 0.5 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 0.5 & -0.5 \end{pmatrix}$$

Now $V(P - I) = 0 \Rightarrow (v_1\ v_2) \begin{pmatrix} -1 & 1 \\ 0.5 & -0.5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$\Rightarrow -v_1 + 0.5 v_2 = 0. \Rightarrow$$

$$v_1 - 0.5 v_2 = 0.$$

and $v_1 + v_2 = 1 \Rightarrow v_2 = 1 - v_1$

i.e., $-v_1 + (0.5)(1 - v_1) = 0.$

$$-v_1 + 0.5 - 0.5 v_1 = 0 \Rightarrow (-1 - 0.5) v_1 = -0.5$$

$$\Rightarrow 1.5 v_1 = 0.5$$

$$v_1 = \frac{5}{15} = \frac{1}{3}$$

and $v_2 = \frac{2}{3}$.

The transition prob. matrix is given by

$$P = \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.2 & 0.2 & 0.6 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} \text{ and } P_0 = (0.4, 0.4, 0.2)$$

a) Find the distn after 3 transitions

b) Find the limiting probabilities.

Sol: $P_0 = (0.4 \quad 0.4 \quad 0.2)$

a) Find P^2, P^3.

The distn after 3 transitions is

$= P_0 P^3 = (0.4 \quad 0.4 \quad 0.2)$ ( )

=

b) $P =$

$P^2 =$

$P^3 =$

⋮

and subsequent powers are the same to this problem.