GUDLAVALLERU ENGINEERING COLLEGE

(An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada)

Seshadri Rao Knowledge Village, Gudlavalleru – 521 356.

Department of Civil Engineering

HANDOUT

on

NUMERICAL AND STATISTICAL METHODS

Vision

To provide quality education embedded with knowledge, ethics and advanced skills and preparing students globally competitive to enrich the civil engineering research and practice.

Mission

- To aim at imparting integrated knowledge in basic and applied areas of civil engineering to cater the needs of industry, profession and the society at large.
- To develop faculty and infrastructure making the department a centre of excellence providing knowledge base with ethical values and transforming innovative and extension services to the community and nation.
- To make the department a collaborative hub with leading industries and organizations, promote research and development and combat the challenging problems in civil engineering which leads for sustenance of its excellence.

Programme Educational Objectives:

PEO-1 Graduates of the program will have bright careers in Mechanical Engineering domain and allied areas.

PEO-2 Graduates of the program will have life skills, sense of ethical conduct and social responsibility.

PEO-3 Graduates of the program will continue to learn and update their competencies to face dynamically changing technological environment.

NUMERICAL AND STATISTICAL METHODS

Class & Sem.: III B. Tech. – I Sem. Academic Year: 2019 – 2020
Branch: Mechanical Engineering Credits: 3

1. **Brief history and current developments in the subject area:**
“MATHEMATICS IS THE MOTHER OF ALL SCIENCES”,It is a necessary avenue to scientific knowledge, which opens new vistas of mental activity. A sound knowledge of engineering mathematics is essential for the Modern Engineering student to reach new heights in life. So students need appropriate concepts, which will drive them in attaining goals.

Importance of numerical analysis:

Numerical analysis is the area of mathematics and computer science that creates, analyzes, and implements algorithms for solving numerically the problems of continuous mathematics. Such problems originate generally from real-world applications of algebra, geometry, and calculus, and they involve variables which vary continuously. These problems occur throughout the natural sciences, social sciences, medicine, engineering, and business.

Importance of statistics:

In recent decades, the growth of statistics has made itself felt in almost every major phase of human activity, and the most important feature of its growth has been the shift in emphasis from descriptive statistics to the methods of *statistical inference*, or *inductive statistics.* Statistical Inference concerns generalizations based on sample data; it applies to such problems as *estimating* an engine’s average emission of pollutants on the basis of some trial runs, *testing* a manufacture’s claim on the basis of measurements performed on samples of a product, and predicting the fidelity of an audio system on the basis of sample data pertaining to the performance of its components.

We shall approach the subject of statistics as a science, developing each statistical idea insofar as possible from its probabilistic foundation, and applying each idea to problems of physical or engineering science as soon as it has been developed. There are few areas where the impact of the recent growth of statistics has been felt more strongly than in engineering and industrial management.

2. **Pre-Requisites:**
(a) Basic Knowledge of Mathematics at Intermediate Level is required.
(b) Fundamental definitions of probability beginning from sample space.

3. **Course Objectives:**
(a) To understand the various numerical techniques.
(b) To understand the concepts of probability and statistics.
(c) To know the importance of correlation coefficient & lines of regression.
(d) To know sampling theory and principles of hypothesis testing.

4. **Course Outcomes:**
Students should be able to:
CO1: apply numerical techniques for algebraic, transcendental and ordinary differential equations.
CO2: find polynomial for unequal intervals using Lagrange’s interpolation.
CO3: compute probabilities in different situations.
CO4: use probability distribution in appropriate scenario.
CO5: measure of correlation between variables and to obtain lines of regression.
CO6: construct sampling distribution and calculate its mean and standard deviation.
CO7: apply the appropriate tests to give valid inference.

Program Outcomes:

a. Apply knowledge of Mathematics, Science and Engineering to solve complex Mechanical Engineering problems.
b. Identify, formulate and analyze problems related to Mechanical Engineering.
c. Design Mechanical engineering systems, to meet the desired needs with the economic, environmental, social, ethical, health and safety constraints.
d Investigate the technological challenges through the use of research based knowledge to design experiments, critical analysis and interpretation of data, synthesis of the data to arrive at valid conclusions.
e. Model and simulate Mechanical engineering systems, to conduct experiments and analyze the performance using modern software tools.
f. Assesses issues pertaining to societal, health, safety, legal, cultural and accordingly engage in professional engineering practices.
g. Demonstrate knowledge for sustainable development with an understanding on the impact of professional engineering on society and environment.
h. Follow professional ethics, norms and standards of engineering practices.
i. Work as an effective member of the team and also as an individual in diverse and multi disciplinary streams.
j. Prepare reports and present effectively and also orally communicate fluently with the society and engineering community.
k. Apply knowledge of Management and Finance for effective project management.
l. Engage in life-long learning independently to stay with the changes in technology.

6. Mapping of Course Outcomes with Program Outcomes:

	a	b	c	d	e	f	g	h	i	j	k	l
CO1	M	L										
CO2	L	L										
CO3	M	L										
CO4	M	M										
CO5	M	L										
CO6	M	**L**										
CO7	H	**M**										

7. Prescribed Text Books:

(a) Title: HigherEngineeringMathematics
Author: B.S.Grewal. Publishers: Khanna Publishers, 42nd edition,2012 , New Delhi.
(b) Title: HigherEngineeringMathematics
Author:B.V.Ramana. Publishers: Tata-McGraw Hill company Ltd.
(c) Title: Probability and Statistics
Authors: Dr. T.K.V. Iyengar, Dr. B. Krishna Gandhi, S. Ranganatham, Dr. M.V.S.S.N. Prasad.Publishers: S. Chand & Company Ltd.
(d) Title: Probability and Statistics for Engineers
Authors: Miller, John E. Freund. Publishers: Prentice Hall of India (PHI)

8.Reference Books:

(a) Title: Advanced Engineering Mathematics
Author: Erwin Kreyszig.Publishers:MaitreyPrintech Pvt. Ltd, 8th edition, 2009, Noida.
(b) Title: Fundamentals of Mathematical Statistics
Authors: S.C. Gupta & V.K. Kapoor. Publishers: S. Chand & Company Ltd.
(c) Title: Probability, Statistics and Queuing Theory Applications for Computer Sciences
Authors: Trivedy, Publishers: John Wiley (2nd edition)

9. URLs and Other e – Learnings:

a. www.mathworld .wolfram.com
b. www.social research methods. net/kb/samprob. php
c. www.fourmilabch/rpkp/experiments/statistics.html
d. www.Hypothesis –Testing.html
e. http://quizlet.com
f. www.probabilitycourse. com

10. Digital Learning Material:

- http://nptel.ac.in/courses/106106094
- http://nptel.ac.in/courses/106106094/40
- http://www.socr.ucla.edu
- www.statlect.com
- www.stat.ucla.edu

11. Lecture Schedule:

<table>
<tr><th rowspan="2">Topic</th><th colspan="2">No. of Periods</th></tr>
<tr><th>Theory</th><th>Tutorial</th></tr>
<tr><td colspan="3">UNIT –1:Algebraic and Transcendental Equations</td></tr>
<tr><td>Algebraic and Transcendental Equations-Introduction</td><td>1</td><td rowspan="2">1</td></tr>
<tr><td>Bisection method</td><td>2</td></tr>
<tr><td>Method of False position</td><td>2</td><td rowspan="2">1</td></tr>
<tr><td>Newton-Raphson Method</td><td>1</td></tr>
<tr><td colspan="3">UNIT – 2: Interpolation</td></tr>
<tr><td>Introduction</td><td>1</td><td rowspan="3">1</td></tr>
<tr><td>Finite differences-Forward, Backward and Central Difference Operators</td><td>2</td></tr>
<tr><td>Relation between Operators</td><td>2</td></tr>
<tr><td>Newton’s formulae for Interpolation</td><td>2</td><td rowspan="2">1</td></tr>
<tr><td>Lagrange’s Interpolation</td><td>1</td></tr>
<tr><td colspan="3">UNIT – 3: Numerical Solution of 1st Order Ordinary Differential Equations</td></tr>
<tr><td>Introduction</td><td>1</td><td rowspan="2">1</td></tr>
<tr><td>Taylor’s Series</td><td>2</td></tr>
<tr><td>Euler’s Method</td><td>2</td><td rowspan="3">1</td></tr>
<tr><td>Modified Euler’s Method</td><td>1</td></tr>
<tr><td>Fourth Order Runge-Kutta Method</td><td>1</td></tr>
<tr><td colspan="3">UNIT – 4: Probability and Expectation of Random Variables</td></tr>
<tr><td>Introduction to Probability</td><td>1</td><td rowspan="2">1</td></tr>
<tr><td>Additive, Conditional and Multiplicative Laws of probability</td><td>2</td></tr>
<tr><td>Baye’s Theorem</td><td>2</td><td rowspan="3">1</td></tr>
<tr><td>Discrete Distribution Function-Mean, Variance</td><td>1</td></tr>
<tr><td>Continuous Distribution Function-Mean, Variance</td><td>1</td></tr>
<tr><td colspan="3">UNIT – 5: Probability Distributions and Correlation &Regression</td></tr>
<tr><td>Binomial and Poisson Distribution</td><td>2</td><td rowspan="2">1</td></tr>
<tr><td>Normal Distribution and its properties</td><td>2</td></tr>
<tr><td>Uniform and Exponential Distribution (Mean, Variance)</td><td>2</td><td></td></tr>
<tr><td>Simple Correlation</td><td>1</td><td rowspan="2">1</td></tr>
<tr><td>Linear Regression Lines</td><td>1</td></tr>
<tr><td colspan="3">UNIT – 6: Sampling and Statistical Inference</td></tr>
<tr><td>Introduction on Sampling and Sampling techniques</td><td>1</td><td>1</td></tr>
</table>

<table>
<tr><td>Sampling Distribution (With and Without Replacement)</td><td>2</td><td></td></tr>
<tr><td>Introduction to Statistical Inference</td><td>1</td><td rowspan="2">1</td></tr>
<tr><td>Test for Means (One Sample and Two Samples When the Sample Size is Large)</td><td>2</td></tr>
<tr><td>Test for Proportions (One sample and Two Samples When the Sample Size is Large)</td><td>2</td><td rowspan="2">1</td></tr>
<tr><td>Chi-Square Test (Goodness of Fit)</td><td>1</td></tr>
<tr><td>F-Test (Test for Population Variances)</td><td>1</td><td rowspan="2">1</td></tr>
<tr><td>Introduction to t-test</td><td>1</td></tr>
<tr><td>TOTAL</td><td>47</td><td>14</td></tr>
</table>

12. Seminar Topics

- Lagrange's Interpolation
- Fourth order Runge-Kutta Method
- Probability
- Sampling Distributions
- Significance Tests
- Correlation and Regression

NUMERICAL AND STATISTICAL METHODS

UNIT-I

Algebraic and Transcendental Equations

Learning Material

Course Objectives:

Student should be able to

- Know about the algebraic and Transcendental Equations.
- Understand the Bisection method , method of False Position and Newton -Raphson Method.

Syllabus:

Solution of Algebraic and Transcendental Equations- Introduction – Bisection Method – Method of False Position Method – Newton-Raphson Method.

Learning Outcomes:

Students will be able to

- Find an approximate solution to Algebraic and Transcendental equations using Numerical Methods (with the aid of calculator)

Algebraic and Transcendental equations

Introduction: A problem of great importance in science and engineering is that of determining the roots/ zeros of an equation of the form f(x) = 0

- Polynomial function: A function f(x) is said to be a polynomial function if f(x) is a polynomial in x.
  i.e. f(x) $= a_0x^n + a_1x^{n-1} + ... + a^{n-1}x + a_n$ where $a_0 \neq 0$, the coefficients
$a_0, a_1 ....a_n$ are real constants and n is a non-negative integer.
- Algebraic function: A function which is a sum (or) difference (or) product of two polynomials is called an algebraic function. Otherwise, the function is called a transcendental (or) non-algebraic function.

  Eg: $f(x) = c_1e^x + c_2e^{-x}$

  $f(x) = e^{5x} - \frac{x^3}{2} + 3$
- Algebraic Equation: If f(x) is an algebraic function, then the equation f(x) =0 is called an algebraic equation.
- Transcendental Equation: An equation which contains polynomials,

exponential functions, logarithmic functions and Trigonometric functions etc. is called a Transcendental equation.

Ex:- $xe^{2x} - 1 = 0$, $\cos x - x e^{x} = 0$, $\tan x = x$ are transcendental equations.

- Root of an equation: A number α is called a root of an equation f(x) = 0 if f(α) =0.

  we also say that α is a zero of the function.

  Note: (1) The roots of an equation are the abscissas of the points where the graph y = f(x) cuts the x-axis.

  (2) A polynomial equation of degree n will have exactly n roots, Real or complex, simple or multiple. A transcendental equation may have one root or infinite number of roots depending on the form of f(x).

Methods for solving the equation

•Direct method:

We know the solution of the polynomial equations such as linear equation $ax+b=0$ and quadratic equation $ax^2+bx+c=0$, will be obtained using direct methods or analytical methods. Analytical methods for the solution of cubic and quadratic equations are also well known to us .

There are no direct methods for solving higher degree algebraic equations or equations involving transcendental functions. Such equations are solved by numerical methods.

In these methods we find an interval in which the root lies.

We use Intermediate value theorem of calculus to determine the interval in which the real root of the equation exists.

•Intermediate value theorem: If f(x) is continuous function in the interval [a, b] and f(a)f(b)< 0, then there exists at least one real root 'c' in the interval (a, b) such that f(c) = 0.

In this unit we will study some important methods of solving algebraic and transcendental equations.

•Bisection method: Bisection method is a simple iteration method to solve an equation. This method is also known as" Bolzano method" or "Interval-Halving method". Suppose an equation of the form $f(x)=0$ has exactly one real root between two real numbers x_0, x_1. The numbers are chosen such that $f(x_0)$ and $f(x_1)$ will have opposite signs. Let us bisect the interval $[x_0, x_1]$ and midpoint $x_2 = \frac{x_0 + x_1}{2}$. If $f(x_2)=0$ then x_2 is a root.

If $f(x_1)$ and $f(x_2)$ have same sign then the root lies between x_0 and x_2. The interval is taken as (x_0, x_2) Otherwise the root lies in the interval $[x_2, x_1]$.

Repeating the process of bisection , we obtain successive subintervals which are smaller. At each iteration, we get the mid-point as a better approximation of the root. This process is terminated when interval is smaller than the desired accuracy.

Problems:-1) Find a root of the equation $x^3-5x+1=0$ using the bisection method in 5 – stages

Sol: Let $f(x)=x^3-5x+1$

we note that f(0) > 0 and f(1)< 0

$\therefore$ Root lies between 0 and 1

Consider $x_0=0$ *and* $x_1=1$

By bisection method the next approximation is

$$x_2=\frac{x_0+x_1}{2}=\frac{1}{2}(0+1)=0.5$$

$$\Rightarrow f(x_2)=f(0:5)=-1.375<0 \text{ and } f(0)>0$$

We have the root lies between 0 and 0.5

Now $x_3=\frac{0+0.5}{2}=0.25$

We find $f(x_3)=-0.234375<0$ *and* $f(0)>0$

Since $f(0)>0$, we conclude that root lies between x_0 *and* x_3

The third approximation of the root is

$$x_4=\frac{x_0+x_3}{4}=\frac{1}{2}(0+0.25)$$

$$=0.125$$

We have $f(x_4)=0.37495>0$

Since $f(x_4)>0$ *and* $f(x_3)<0$, the root lies between

$x_4=0.125$ *and* $x_3=0.25$

Considering the 4th approximation of the roots

$$x_5=\frac{x_3+x_4}{2}=\frac{1}{2}(0.125+0.25)=0.1875$$

$f(x_5)=0.06910>0$,

since $f(x_5)>0$ *and* $f(x_3)<0$ the root must lie between $x_5=0.18758$ and $x_3=0.25$

Here the fifth approximation of the root is

$$x_6=\frac{1}{2}(x_5+x_3)$$

$$=\frac{1}{2}(0.1875+0.25)$$

$$=0.21875$$

We are asked to do up to 5 stages.
We stop here 0.21875 is taken as an approximate value of the root and it lies between 0 and 1

False Position Method (Regula – Falsi Method)

In the false position method we will find the root of the equation $f(x)=0$.
Consider two initial approximate values x_0 *and* x_1 near the required root so that $f(x_0)$ *and* $f(x_1)$ have different signs. This implies that a root lies between x_0 *and* x_1. The curve $f(x)$ crosses x- axis only once at the point x_2 lying between the points x_0 *and* x_1, Consider the point $A=(x_0, f(x_0))$ and $B=(x_1, f(x_1))$ on the graph and suppose they are connected by a straight line, Suppose this line cuts x-axis at x_2, We calculate the values of $f(x_2)$ at the point. If $f(x_0)$ *and* $f(x_2)$ are of opposite sign, then the root lies between x_0 *and* x_2 and value x_1 is replaced by x_2

Otherwise the root lies between x_2 and x_1 and the value of x_0 is replaced by x_2

Another line is drawn by connecting the newly obtained pair of values. Again the point here the line cuts the x-axis is a closer approximation to the root. This process is repeated as many times as required to obtain the desired accuracy. It can be observed that the points x_2, x_3, x_4 .....obtained converge to the expected root of the equation $y=f(x)$.

To obtain the equation to find the next approximation to the root

Let $A=(x_0, f(x_0))$ *and* $B=(x_1, f(x_1))$ be the points on the curve $y=f(x)$ Then the equation to the chord AB is $\frac{y-f(x_0)}{x-x_0}=\frac{f(x_1)-f(x_0)}{x-x_0}\rightarrow(1)$

At the point C where the line AB crosses the x – axis, we have f(x) =0 i.e. y =0

From (1), we get $x=x_0-\frac{x_1-x_0}{f(x_1)-f(x_0)}f(x_0)\rightarrow(2)$

x is given by (2) serves as an approximated value of the root, when the interval in which it lies is small. If the new values of x is taken as x_2 then (2) becomes

$$x_2=x_0-\frac{(x_1-x_0)}{f(x_1)-f(x_0)}f(x_0)$$

$$=\frac{x_0 f(x_1)-x_1 f(x_0)}{f(x_1)-f(x_0)} \text{ ---------------3}$$

Now we decide whether the root lies between x_0 *and* x_2 $(or)\, x_2$ *and* x_1

We name that interval as (x_1, x_2) The line joining $(x_1, y_1)(x_2, y_2)$ meets x – axis at x_3 is given by $x_3 = \dfrac{x_1 f(x_2) - x_2 f(x_1)}{f(x_2) - f(x_1)}$

This will in general, be nearest to the exact root we continue this procedure till the root is found to the desired accuracy. The iteration process based on (3) is known as the method of false position. The successive intervals where the root lies, in the above procedure are named as $(x_0, x_1), (x_1, x_2), (x_2, x_3)$ etc

Where $x_i < x_i + 1$ and $f(x_0), f(x_i + 1)$ are of opposite signs

Also $x_{i+1} = \dfrac{x_{i-1} f(x_i) - x_i f(x_{i-1})}{f(x_i) - f(x_{i-1})}$

Problems:-

1.Find out the roots of the equation $x^3 - x - 4 = 0$ using false position method

sol: Let $f(x) = x^3 - x - 4 = 0$

$f(0) = -4, f(1) = -4, f(2) = 2$

Since $f(1)$ *and* $f(2)$ have opposite signs the root lies between 1 and 2

By false position method $x_2 = \dfrac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)}$

$$x_2 = \frac{(1 \times 2) - 2(-4)}{2 - (-4)}$$

$$= \frac{2+8}{6} = \frac{10}{6} = 1.666$$

$$f(1.666) = (1.666)^3 - 1.666 - 4$$
$$= -1.042$$

Now, the root lies between 1.666 and 2

$$x_3 = \frac{1.666 \times 2 - 2 \times (-1.042)}{2 - (-1.042)} = 1.780$$

$$f(1.780) = (1.780)^3 - 1.780 - 4$$
$$= -0.1402$$

Now, the root lies between 1.780 and 2

$$x_4 = \frac{1.780 \times 2 - 2 \times (-0.1402)}{2 - (-0.1402)} = 1.794$$

$$f(1.794) = (1.794)^3 - 1.794 - 4$$
$$= -0.0201$$

Now, the root lies between 1.794 and 2

$$x_5 = \frac{1.794 \times 2 - 2 \times (-0.0201)}{2 - (-0.0201)} = 1.796$$

$$f(1.796) = (1.796)^3 - 1.796 - 4 = -0.0027$$

Now, the root lies between 1.796 and 2

$$x_6 = \frac{1.796 \times 2 - 2 \times (-0.0027)}{2 - (-0.0027)} = 1.796$$

The root is 1.796.

<u>Newton- Raphson Method:-</u>

The Newton- Raphson method is a powerful and elegant method to find the root of an equation. This method is generally used to improve the results obtained by the previous methods.

Let x_0 be an approximate root of $f(x) = 0$ and let $x_1 = x_0 + h$ be the correct root which implies that $f(x_1) = 0$.

By Taylor's theorem neglecting second and higher order terms

$$f(x_1) = f(x_0 + h) = 0$$

$$\Rightarrow f(x_0) + hf^1(x_0) = 0$$

$$\Rightarrow h = -\frac{f(x_0)}{f^1(x_0)}$$

Substituting this in x_1 we get

$$x_1 = x_0 + h$$

$$= x_0 - \frac{f(x_0)}{f^1(x_0)}$$

$\therefore x_1$ is a better approximation than x_0

Successive approximations are given by

$$x_{n+1} = x_n - \frac{f(x_n)}{f^1(x_n)}$$

Problem:- **1**.Find by Newton's method, the real root of the equation $xe^x - 2 = 0$ Correct to three decimal places.

Sol. Let $f(x) = xe^x - 2 \rightarrow (1)$

Then $f(0) = -2$ and $f(1) = e - 2 = 0.7183$

So root of $f(x)$ lies between 0 and 1

It is near to 1. so we take $x_0 = 1$ and $f^1(x) = xe^x + e^x$ *and* $f^1(1) = e + e = 5.4366$

$\therefore$ By Newton's Rule

First approximation $x_1 = x_0 - \frac{f(x_0)}{f^1(x_0)}$

$$= 1 - \frac{0.7183}{5.4366} = 0.8679$$

$\therefore f(x_1) = 0.0672 \qquad f^1(x_1) = 4.4491$

The second approximation $x_2 = x_1 - \frac{f(x_1)}{f^1(x_1)}$

$$= 0.8679 - \frac{0.0672}{4.4491}$$

$$= 0.8528$$

$\therefore$ Required root is 0.853 correct to 3 decimal places.

Convergence of the Iteration Methods

We now study the rate at which the iteration methods converge to the exact root, if the initial approximation is sufficiently close to the desired root.

Define the error of approximation at the k th iterate as $\epsilon_k = x_k - \alpha$, k = 0, 1, 2,...
Definition: An iterative method is said to be of order p or has the rate of convergence p, if p is the largest positive real number for which there exists a finite constant C≠0 , such that

$$|\epsilon_{k+1}| < | \epsilon_k^p |$$

The constant C, which is independent of k, is called the asymptotic error constant and it depends on the derivatives of f(x) at x = α.

NUMERICAL AND STATISTICAL METHODS
Learning Material

UNIT-II

INTERPOLATION

Objectives:

- Develop an understanding of the use of numerical methods in modern scientific computing.
- To gain the knowledge of Interpolation.

Syllabus:

Interpolation- Introduction – Finite differences- Forward Differences – Backward differences –Central differences – Symbolic relations – Newton formulae for interpolation – Lagrange's interpolation.

Learning Outcomes:

Student should be able to

- Know about the Interpolation, and Finite Differences.
- Utilize the Newton's formula for interpolation.
- Operate Lagrange's Interpolation formula.

Introduction:-

Consider the eqation y =f(x) , $x_0 \leq x \leq x_n$ we understand that we can find the value of y, corresponding to every value of x in the range $x_0 \leq x \leq x_n$. If the function f(x) is single valued and continuous and is known explicitly, then the values of f(x) for certain values of x like $x_0, x_1, \ldots\ldots, x_n$ can be calculated. The problem now is if we are given the set of tabular values

$$x : x_0 \quad x_1 \quad x_2 \ldots\ldots x_n$$

$$y : y_0 \quad y_1 \quad y_2 \ldots\ldots y_n$$

Satisfying the relation y = f(x) and the explicit definition of f(x) is not known, is it possible to find a simple function say f(x) such that f(x) and $\phi(x)$ agree at the set of tabulated points. This process of finding $\phi(x)$ is called interpolation. If $\phi(x)$ is a polynomial then the process is called polynomial interpolation and $\phi(x)$ is called interpolating polynomial. In our study we are concerned with polynomial interpolation

Finite Differences:-

1. **Introduction**:- Here we introduce forward, backward and central differences of a function y = f(x). These differences play a fundamental role in the study of differential calculus, which is an essential part of numerical applied mathematics

2. Forward Differences:-

Consider a function y = f(x) of an independent variable x. let $y_0, y_1, y_2 \ldots\ldots y_r$ be the values of y corresponding to the values $x_0, x_1, x_2 \ldots\ldots x_r$ of x respectively. Then the differences $y_1 - y_0$, $y_2 - y_1 \ldots\ldots\ldots\ldots$ are called the first forward differences of y, and we denote them by $\Delta y_0, \Delta y_1, \ldots\ldots$ that is

$$\Delta y_0 = y_1 - y_0, \Delta y_1 = y_2 - y_1, \Delta y_2 = y_3 - y_2 \ldots\ldots$$

In general $\Delta y_r = y_{r+1} - y_r \therefore r = 0,1,2-----$

Here the symbol Δ is called the forward difference operator

The second forward differences and are denoted by $\Delta^2 y_0, \Delta^2 y_1 ......$ that is

$$\Delta^2 y_0 = \Delta y_1 - \Delta y_0$$

$$\Delta^2 y_1 = \Delta y_2 - \Delta y_1$$

In general $\Delta^2 y_r = \Delta y_{r+1} - \Delta y_r \quad r = 0,1,2.......$ similarly, the nth forward differences are defined by the formula.

$$\Delta^n y_r = \Delta^{n-1} y_{r+1} - \Delta^{n-1} y_r \quad r = 0,1,2.......$$

The symbol Δ^n is referred as the nth forward difference operator.

3. Forward Difference Table:-

The forward differences are usually arranged in tabular columns as shown in the following table called a forward difference table

Values of x	Values of y	First order differences	Second order differences	Third order differences	Fourth order differences
x_o	y_0				
		$\Delta y_0 = y_1 - y_0$			
x_1	y_1		$\Delta^2 y_0 = \Delta y_1 - y_0$		
		$\Delta y_1 = y_2 - y_1$		$\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0$	
x_2	y_2		$\Delta^2 y_1 = \Delta y_2 - \Delta y_1$		$\Delta^4 y_0 = \Delta^3 y_1 - \Delta^3 y_0$
		$\Delta y_2 = y_3 - y_2$		$\Delta^3 y_1 = \Delta^2 y_2 - \Delta^2 y_1$	
x_3	y_3		$\Delta^2 y_2 = \Delta y_3 - \Delta y_2$		
x_4	y_4	$\Delta y_3 = y_4 - y_3$			

4. Backward Differences:-

Let $y_0, y_1 ...... y_r ......$ be the values of a function $y = f(x)$ corresponding to the values $x_0, x_1, x_2 .... x_r ......$ of x respectively. Then, $\nabla y_1 = y_1 - y_0, \nabla y_2 = y_2 - y_1, \nabla y_3 = y_3 - y_2, ...$ are called the first backward differences

In general $\nabla y_r = y_r - y_{r-1}, \ r = 1,2,3......... \to (1)$

The symbol ∇ is called the backward difference operator, like the operator Δ, this operator is also a linear operator

Comparing expression (1) above with the expression (1) of section we immediately note that $\nabla y_r = \nabla y_{r-1}, r = 0,1,2....... \to (2)$

The first backward differences of the first background differences are called second differences and are denoted by $\nabla^2 y_2, \nabla^2 y_3 ---\nabla^2_r ----$ i.e.,..

$$\nabla^2 y_2 = \nabla y_2 - \nabla y_1, \nabla^2 y_3 = \nabla y_3 - \nabla y_2 ....$$

In general $\nabla^2 y_r = \nabla y_r - \nabla y_{r-1}, r = 2,3..... \to (3)$ similarly, the nth backward differences are defined by the formula $\nabla^n y_r = \nabla^{n-1} y_r - \nabla^{n-1} y_{r-1}, r = n, n+1..... \to (4)$

If $y = f(x)$ is a constant function , then y = c is a constant, for all x, and we get $\nabla^n y_r = 0 \forall n$ the symbol ∇^n is referred to as the n^th backward difference operator

5. Backward Difference Table:-

X	Y	∇y	$\nabla^2 y$	$\nabla^3 y$
x_0	y_0			
		∇y_1		
x_1	y_1		$\nabla^2 y_2$	
		∇y_2		$\nabla^3 y_3$
x_2	y_2		$\nabla^2 y_3$	
		∇y_3		
x_3	y_3			

6. Central Differences:-

With $y_0, y_1, y_2 .... y_r$ as the values of a function $y = f(x)$ corresponding to the values $x_1, x_2 ..... x_r .... \text{ of } x$, we define the first central differences

$\delta y_{1/2}, \delta y_{3/2}, \delta y_{5/2} ----$ as follows

$$\delta y_{1/2} = y_1 - y_0, \delta y_{3/2} = y_2 - y_1, \delta y_{5/2} = y_3 - y_2 ----$$

$$\delta y_{r-1/2} = y_r - y_{r-1} \rightarrow (1)$$

The symbol δ is called the central differences operator. This operator is a linear operator. Comparing expressions (1) above with expressions earlier used on forward and backward differences we get

$\delta y_{1/2} = \Delta y_0 = \nabla y_1, \delta y_{3/2} = \Delta y_1 = \nabla y_2 .....$

In general $\delta y_{n+1/2} = \Delta y_n = \nabla y_{n+1}, n = 0,1,2 ...... \rightarrow (2)$

The first central differences of the first central differences are called the second central differences and are denoted by $\delta^2 y_1, \delta^2 y_2 ...$

Thus $\delta^2 y_1 = \delta_{3/2} - \delta y_{1/2}, \delta^2 y_2 = \delta_{5/2} - \delta_{3/2} ....$

$$\delta^2 y_n = \delta y_{n+1/2} - \delta y_{n-1/2} \rightarrow (3)$$

Higher order central differences are similarly defined. In general the n^th central differences are given by

for odd $n : \delta^n y_{r-1/2} = \delta^{n-1} y_r - \delta^{n-1} y_{r-1}, r = 1,2 .... \rightarrow (4)$

for even $n : \delta^n y_r = \delta^{n-1} y_{r+1/2} - \delta^{n-1} y_{r-1/2}, r = 1,2 .... \rightarrow (5)$

while employing for formula (4) for $n = 1$, we use the notation $\delta^0 y_r = y_r$

If y is a constant function, that is if $y = c$ a constant, then $\delta^n y_r = 0 \;\; for \;\; all \;\; n \geq 1$

7. Central Difference Table

x_0	y_0	δy	$\delta^2 y$	$\delta^3 y$	$\delta^4 y$
		$\delta y_{1/2}$			
x_1	y_1		$\delta^2 y_1$		
		$\delta y_{2/2}$		$\delta^3 y_{3/2}$	
x_2	y_2		$\delta^2 y_2$		$\delta^4 y_2$
		$\delta y_{5/2}$		$\delta^3 y_{5/2}$	
x_3	y_3		$\delta^2 y_3$		
		$\delta y_{7/2}$			
x_4	y_4				

Symbolic Relations :

E-operator:- The shift operator E is defined by the equation $Ey_r = y_{r+1}$. This shows that the effect of E is to shift the functional value y_r to the next higher value y_{r+1}. A second operation with E gives $E^2 y_r = E(Ey_r) = E(y_{r+1}) = y_{r+2}$

Generalizing $E^n y^r = y_{r+n}$

Averaging operator:- The averaging operator μ is defined by the equation $\mu y_r = \frac{1}{2}[y_{r+1/2} + y_{r-1/2}]$

Relationship Between Δ *and* E

We have

$$\Delta y_0 = y_1 - y_0$$
$$= Ey_0 - y_0 = (E-1)y_0$$
$$\Rightarrow \Delta = E - y(or) E = 1 + \Delta$$

Some more relations

$$\Delta^3 y_0 = (E-1)^3 y_0 = (E^3 - 3E^2 + 3E - 1)y_0$$
$$= y_3 - 3y_2 + 3y_1 - y_0$$

•**Inverse operator**: Inverse operator E^{-1} is defined as $E^{-1}y_r = y_{r-1}$

In general $E^{-n}y_n = y_{r-n}$

We can easily establish the following relations

i) $\nabla \equiv 1 - E^{-1}$ ii) $\delta \equiv E^{1/2} - E^{-1/2}$ iii) $\mu = \frac{1}{2}(E^{1/2} + E^{-1/2})$

iv) $\Delta = \nabla E = E^{1/2}$ v) $\mu^2 \equiv 1 + \frac{1}{4}\delta^2$

•**Differential operator:**

The operator D is defined as $Dy(x) = \frac{\partial}{\partial x}[y(x)]$

Relation between the Operators D and E

Using Taylor's series we have, $y(x+h)=y(x)+hy^{1}(x)+\frac{h^2}{2!}y^{11}(x)+\frac{h^3}{3!}y^{111}(x)+----$

This can be written in symbolic form

$$Ey_x=\left[1+hD+\frac{h^2D^2}{2!}+\frac{h^3D^3}{3!}+----\right]y_x \;=e^{hD}.y_x$$

We obtain in the relation $E=e^{hD}\rightarrow(3)$

•**Theorem:** If $f(x)$ is a polynomial of degree n and the values of x are equally spaced then $\Delta^n f(x)$ is constant

Note:-

As $\Delta^n f(x)$ is a constant, it follows that $\Delta^{n+1}f(x)=0,\Delta^{n+2}f(x)=0,........$

The converse of above result is also true that is, if $\Delta^n f(x)$ is tabulated at equal spaced intervals and is a constant, then the function $f(x)$ is a polynomial of degree n

1. Find the missing term in the following data

X	0	1	2	3	4
Y	1	3	9	-	81

Why this value is not equal to 3^3. Explain

Sol. Consider $\Delta^4 y_0=0$

$\Rightarrow 4y_0-4y_3+5y_2-4y_1+y_0=0$

Substitute given values we get

$81-4y_3+54-12+1=0\Rightarrow y_3=31$

From the given data we can conclude that the given function is $y=3^x$. To find y_3, we have to assume that y is a polynomial function, which is not so. Thus we are not getting $y=3^3=27$.

2. Evaluate

$(i)\Delta\cos x$

$(ii)\Delta^2\sin(px+q)$

$(iii)\Delta^n e^{ax+b}$

Sol. Let h be the interval of differencing

$$(i)\Delta\cos x=\cos(x+h)-\cos x$$

$$=-2\sin\left(x+\frac{h}{2}\right)\sin\frac{h}{2}$$

$$(ii)\Delta\sin(px+q)=\sin\left[p(x+h)+q\right]-\sin(px+q)$$

$$=2\cos\left(px+q+\frac{ph}{2}\right)\sin\frac{ph}{2}$$

$$=2\sin\frac{ph}{2}\sin\left(\frac{\pi}{2}+px+q+\frac{ph}{2}\right)$$

$$\Delta^2 \sin(px+q) = 2\sin\frac{ph}{2}\Delta\left[\sin(px+q)+\frac{1}{2}(\pi+ph)\right]$$
$$=\left[2\sin\frac{ph}{2}\right]^2 \sin\left[px+q+\frac{1}{2}(\pi+ph)\right]$$

$$(iii)\ \Delta e^{ax+b} = e^{a(x+h)+b} - e^{ax+b}$$
$$= e^{(ax+b)}\left(e^{ah-1}\right)$$
$$\Delta^2 e^{ax+b} = \Delta\left[\Delta\left(e^{ax+b}\right)\right] - \Delta\left[\left(e^{ah}-1\right)\left(e^{ax+b}\right)\right]$$
$$=\left(e^{ah}-1\right)^2 \Delta\left(e^{ax+h}\right)$$
$$=\left(e^{ah}-1\right)^2 e^{ax+b}$$

Proceeding on, we get $\Delta^n\left(e^{ax+b}\right) = \left(e^{ah}-1\right)^n e^{ax+b}$

Newton's Forward Interpolation Formula:-

Let $y=f(x)$ be a polynomial of degree n and taken in the following form

$$y=f(x)=b_0+b_1(x-x_0)+b_2(x-x_0)(x-x_1)+b_3(x-x_0)(x-x_1)(x-x_2)+---$$
$$+b_n(x-x_0)(x-x_1)----(x-x_{n-1})\to(1)$$

This polynomial passes through all the points (for i = 0 to n. Therefore, we can obtain the $y_i's$ by substituting the corresponding $x_i's$ as

$$at\ \ x=x_0, y_0=b_0$$
$$at\ \ x=x_1, y_1=b_0+b_1(x_1-x_0)$$
$$at\ \ x=x_2, y_2=b_0+b_1(x_2-x_0)+b_2(x_2-x_0)(x_2-x_1)\to(1)$$

Let 'h' be the length of interval such that $x_i's$ represent

$$x_0, x_0+h, x_0+2h, x_0+3h----x_0+xh$$

This implies $x_1-x_0=h, x_2-x_0-2h, x_3-x_0=3h----x_n-x_0=nh\to(2)$

From (1) and (2), we get

$$y_0=b_0$$
$$y_1=b_0+b_1h$$
$$y_2=b_0+b_12h+b_2(2h)h$$
$$y_3=b_0+b_13h+b_2(3h)(2h)+b_3(3h)(2h)h$$

....

....

$$y_n=b_0+b_1(nh)+b_2(nh)(n-1)h+----+b_n(nh)\left[(n-1)h\right]\left[(n-2)h\right]\to(3)$$

Solving the above equations for $b_0, b_{11}, b_2.....b_n$, we get $b_0=y_0$

$$b_1 = \frac{y_1 - b_0}{h} = \frac{y_1 - y_0}{h} = \frac{\Delta y_0}{h}$$

$$b_2 = \frac{y_2 - b_0 - b_1 2h}{2h^2} = y_2 - y_0 - \frac{(y_1 - y_0)}{h} 2h$$

$$= \frac{y_2 - y_0 - 2y_1 - 2y_0}{2h^2} = \frac{y_2 - 2y_1 + y_0}{2h^2} = \frac{\Delta^2 y_0}{2h^2}$$

$$\therefore b_2 = \frac{\Delta^2 y_0}{2!h^2}$$

Similarly, we can see that

$$b_3 = \frac{\Delta^3 y_0}{3!h^3}, b_4 = \frac{\Delta^4 y_0}{4!h^4} - - - - - b_n = \frac{\Delta^n y_0}{n!h^n}$$

$$\therefore y = f(x) = y_0 + \frac{\Delta y_0}{h}(x - x_0) + \frac{\Delta^2 y_0}{2!h^2}(x - x_0)(x - x_1)$$

$$+ \frac{\Delta^3 y_0}{3!h^3}(x - x_0)(x - x_1)(x - x_2) + - - - +$$

$$+ \frac{\Delta^n y_0}{n!h^n}(x - x_0)(x - x_1) - - - (x - x_{n-1}) \rightarrow (3)$$

If we use the relationship $x = x_0 + ph \Rightarrow x - x_0 = ph,\ where\ p = 0,1,2,.....n$

Then

$$x - x_1 = x - (x_0 + h) = (x - x_0) - h$$

$$= ph - h = (p-1)h$$

$$x - x_2 = x - (x_1 + h) = (x - x_1) - h$$

$$= (p-1)h - h = (p-2)h$$

....

$$x - x_i = (p - i)h$$

....

$$x - x_{n-1} = [p - (n-1)]h$$

Equation (3) becomes

$$y = f(x) = f(x_0 + ph) = y_0 + p\Delta y_0 + \frac{p(p-1)}{2!}\Delta^2 y_0 + \frac{p(p-1)(p-2)}{3!}\Delta^3 y_0 + - - - - +$$

$$\frac{p(p-1)(p-2) - - - - (p - (n-1))}{n!}\Delta^n y_0 \rightarrow (4)$$

Newton's Backward Interpolation Formula:-

If we consider

$$y_n(x) = a_0 + a_1(x - x_n) + a_2(x - x_n)(x - x_{n-1}) + a_3(x - x_n)(x - x_{n-1})(x - x_{n-2}) + - - - - (x - x_i)$$

and impose the condition that y and $y_n(x)$ should agree at the tabulated points $x_n, x_n - 1, ...x_2, x_1, x_0$

We obtain

$$y_n(x) = y_n + p\nabla y_n + \frac{p(p+1)}{2i}\nabla^2 y_n + ---$$
$$\frac{p(p+1)----[p+(n-1)]}{n!}\nabla^n y_n + ---- \rightarrow (6)$$

Where $p = \frac{x - x_n}{h}$

This uses tabular values of the left of y_n. Thus this formula is useful formula is useful for interpolation near the end of the tabular values

Q:-1. Find the melting point of the alloy containing 54% of lead, using appropriate interpolation formula

Percentage of lead(p)	50	60	70	80
Temperature $(Q°c)$	205	225	248	274

Sol. The difference table is

x	Y	Δ	Δ^2	Δ^3
50	205			
		20		
60	225		3	
		23		0
70	248		3	
		26		
80	274			

Let temperature = $f(x)$

$$x_0 + ph = 24, x_0 = 50, h = 10$$
$$50 + p(10) = 54 \ (or)\ p = 0.4$$

By Newton's forward interpolation formula

$$f(x_0 + ph) = y_0 + p\Delta y_0 + \frac{p(p-1)}{2!}\Delta^2 y_0 + \frac{p(p-1)(p-2)}{n!}\Delta^3 y_0 + ---$$
$$f(54) = 205 + 0.4(20) + \frac{0.4(0.4-1)}{2!}(3) + \frac{(0.4)(0.4-1)(0.4-2)}{3!}(0)$$
$$= 205 + 8 - 0.36$$
$$= 212.64$$

Melting point = 212.6

2.Using Newton's Gregory backward formula, find *e* $^{1.9}$ from the following data

x	1.00	1.25	1.50	1.75	2.00
e^{x}	0.3679	0.2865	0.2231	0.1738	0.1353

Lagrange's Interpolation Formula:-

Let $x_0, x_1, x_2, \ldots.\ x_n$ be the $(n+1)$ values of x which are not necessarily equally spaced. Let $y_0, y_1, y_2 \ldots\ldots y_n$ be the corresponding values of $y = f(x)$ let the polynomial of degree n for the function $y = f(x)$ passing through the $(n+1)$ points $(x_0, f(x_0)), (x_1, f(x_1)) ---- (x_n, f(x_n))$ be in the following form

$$y = f(x) = a_0(x-x_1)(x-x_2)\ldots\ldots(x-x_n) + a_1(x-x_0)(x-x_2)\ldots\ldots\ldots(x-x_n) +$$
$$a_2(x-x_0)(x-x_1)\ldots\ldots.(x-x_n) + \ldots\ldots.. + a_n(x-x_0)(x-x_1)\ldots\ldots(x-x_{n-1}) \rightarrow (1)$$

Where $a_0, a_1, a_2 \ldots.$ a^n are constants

Since the polynomial passes through $(x_0, f(x_0))$, $(x_1, f(x_1))\ldots\ldots(x_n, f(x_n))$. The constants can be determined by substituting one of the values of $x_0, x_1, \ldots\ldots x_n$ *for* x in the above equation

Putting $x = x_0$ in (1) we get, $f(x_0) = a_0(x-x_1)(x_0-x_2)(x_0-x_n)$

$$\Rightarrow a_0 = \frac{f(x_0)}{(x-x_1)(x_0-x_2)\ldots.(x_0-x_n)}$$

Putting $x = x_1$ in (1) we get, $f(x_1) = a_1(x-x_0)(x_1-x_2) ---- (x_1-x_n)$

$$\Rightarrow a_1 = \frac{f(x_1)}{(x_1-x_0)(x_1-x_2)\ldots.(x_1-x_n)}$$

Similarly substituting $x = x_2$ in (1), we get

$$\Rightarrow a_2 = \frac{f(x_2)}{(x_2-x_0)(x_2-x_1)\ldots\ldots(x_2-x_n)}$$

Continuing in this manner and putting $x = x_n$ in (1) we get

$$a_n = \frac{f(x_n)}{(x_n-x_0)(x_n-x_1) ---- (x_n-x_{n-1})}$$

Substituting the values of $a_0, a_1, a_2 \ldots.a_n$, we get

$$f(x) = \frac{(x-x_1)(x-x_2)\ldots\ldots.(x-x_n)}{(x_0-x_1)(x_0-x_2)\ldots\ldots\ldots(x_0-x_n)} f(x_0) + \frac{(x-x_0)(x-x_2)\ldots..(x-x_n)}{(x_1-x_0)(x_1-x_2)\ldots.(x_1-x_n)}$$
$$f(x_1) + \frac{(x-x_0)(x-x_1)(x-x_2)\ldots..(x-x_n)}{(x_2-x_0)(x_2-x_1)\ldots\ldots(x_2-x_n)} + \ldots..f(x_2) + \frac{(x-x_0)(x-x_1)\ldots..(x-x_{n-1})}{(x_n-x_1)(x_n-x_2)\ldots..(x_n-x_{n-1})} f(x_n)$$

Q 1. Using Lagrange's formula calculate $f(3)$ from the following table

x	0	1	2	4	5	6
$f(x)$	1	14	15	5	6	19

Sol. Given $x_0 = 0, x_1 = 1, x_2 = 2, x_3 = 4, x_5 = 6, x_4 = 5$

$f(x_0) = 1, f(x_1) = 14, f(x_2) = 15, f(x_3) = 5, f(x_4) = 6, f(x_5) = 19$

From lagrange's interpolation formula

$$f(x) = \frac{(x-x_1)(x-x_2)(x-x_3)(x-x_4)(x-x_5)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)(x_0-x_4)(x_0-x_5)} f(x_0)$$
$$+\frac{(x-x_0)(x-x_2)(x-x_3)(x-x_4)(x-x_5)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)(x_1-x_4)(x_1-x_5)} f(x_1)$$
$$+\frac{(x-x_0)(x-x_1)(x-x_3)(x-x_4)(x-x_5)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)(x_2-x_4)(x_2-x_5)} f(x_2)$$

$$\frac{(x-x_0)(x-x_1)(x-x_2)(x-x_3)(x-x_4)}{(x_5-x_0)(x_5-x_1)(x_5-x_2)(x_5-x_3)(x_5-x_4)} f(x_5)$$

Here $x = 3$ then

$$f(3) = \frac{(3-1)(3-2)(3-4)(3-5)(3-6)}{(0-1)(0-2)(0-4)(0-5)(0-6)} \times 1 +$$
$$\frac{(3-0)(3-2)(3-4)(3-5)(3-6)}{(1-0)(1-2)(1-4)(1-5)(1-6)} \times 14 +$$
$$\frac{(3-0)(3-1)(3-4)(3-5)(3-6)}{(2-0)(2-1)(2-4)(2-5)(2-6)} \times 15 +$$
$$\frac{(3-0)(3-1)(3-2)(3-5)(3-6)}{(4-0)(4-1)(4-2)(4-5)(4-6)} \times 5 +$$
$$\frac{(3-0)(3-1)(3-2)(3-4)(3-6)}{(5-0)(5-1)(5-2)(5-4)(5-6)} \times 6 +$$
$$\frac{(3-0)(3-1)(3-2)(3-4)(3-5)}{(6-0)(6-1)(6-2)(6-4)(6-5)} \times 19$$
$$= \frac{12}{240} - \frac{18}{60} \times 14 + \frac{36}{48} \times 15 + \frac{36}{48} \times 5 - \frac{18}{60} \times 6 + \frac{12}{40} \times 19$$
$$= 0.05 - 4.2 + 11.25 + 3.75 - 1.8 + 0.95$$
$$= 10$$
$$f(x_3) = 10$$

NUMERICAL AND STATISTICAL METHODS

UNIT - III

NUMERICAL SOLUTION OF FIRST ORDER ORDINARY DIFFERENTIAL EQUATIONS

Objectives:

• To the numerical solutions of a first ordered Ordinary Differential Equation together with initial condition.

Syllabus:

Taylor's Series Method - Euler Method - Modified Euler Method - Runge – Kutta Fourth order Method.

Subject Outcomes:

At the end of the unit, Students will be able to

• Solve Ordinary Differential equations using Numerical methods.

The important methods of solving ordinary differential equations of first order numerically are as follows

- Taylors series method
- Euler's method
- Modified Euler's method of successive approximations
- Runge- kutta method

To describe various numerical methods for the solution of ordinary differential equations we consider the general 1st order differential equation

$dy/dx = f(x,y)$ ------- (1)

with the initial condition $y(x_0) = y_0$

The methods will yield the solution in one of the two forms:

i) A series for y in terms of powers of x, ,from which the value of y can be obtained by direct substitution.

ii) A set of tabulated values of y corresponding to different values of x.

TAYLOR'S SERIES METHOD

To find the numerical solution of the differential equation

$$\frac{dy}{dx} = f(x,y) \rightarrow (1)$$

With the initial condition $y(x_0) = y_0 \rightarrow$ (2)

$y(x)$ can be expanded about the point x_0 in a Taylor's series in powers of $(x-x_0)$ as

$$y(x) = y(x_0) + \frac{(x-x_0)}{1} y'(x_0) + \frac{(x-x_0)^2}{2!} y''(x_0) + \dots\dots\dots\dots + \frac{(x-x_0)^n}{n!} y^n(x_0) \rightarrow (3)$$

In equation3, $y(x_0)$ is known from Initial Condition . The remaining coefficients $y'(x_0), y''(x_0), \dots\dots\dots y^n(x_0)$ etc are obtained by successively differentiating equation1 and evaluating at x_0. Substituting these values in equation3, $y(x)$ at any point can be calculated from equation3. Provided $h = x - x_0$ is small.

When $x_0 = 0$, then Taylor's series equ3 can be written as

$$y(x) = y(0) + x.y'(0) + \frac{x^2}{2!} y''(0) + \dots\dots + \frac{x^n}{n!} y^n(0) + \dots\dots\dots \rightarrow (4)$$

Note: We know that the Taylor's expansion of y(x) about the point x_0 in a power of $(x - x_0)$is.

$$y(x) = y(x_0) + \frac{(x-x_0)}{1!} y^{I}(x_0) + \frac{(x-x_0)^2}{2!} y^{II}(x_0) + \frac{(x-x_0)^3}{3!} y^{III}(x_0) + \dots \rightarrow (1)$$

Or

$$y(x) = y_0 + \frac{(x-x_0)}{1!} y_0^I + \frac{(x-x_0)^2}{2!} y_0^{II} + \frac{(x-x_0)^3}{3!} y_0^{III} + \dots\dots$$

If we let $x - x_0 = h$. (i.e. $x = x_0 + h = x_1$) we can write the Taylor's series as

$$y(x) = y(x_1) = y_0 + \frac{h}{1!} y_0^I + \frac{h^2}{2!} y_0^{II} + \frac{h^3}{3!} y_0^{III} + \frac{h^4}{4!} y_0^{IV} + \dots\dots$$

$$\text{i.e. } y_1 = y_0 + \frac{h}{1!} y_0^I + \frac{h^2}{2!} y_0^{II} + \frac{h^3}{3!} y_0^{III} + \frac{h^{IV}}{4!} y_0^{IV} + \dots\dots \rightarrow (2)$$

Similarly expanding y(x) in a Taylor's series about $x = x_1$. We will get.

$$y_2 = y1 + \frac{h}{1!} y_1^I + \frac{h^2}{2!} y_1^{II} + \frac{h^3}{3!} y_1^{III} + \frac{h^4}{4!} y_1^{IV} + \dots\dots\dots \rightarrow (3)$$

Similarly expanding y(x) in a Taylor's series about $x = x_2$ We will get.

$$y_3 = y_2 + \frac{h}{1!} y_2^{I} + \frac{h^2}{2!} y_2^{II} + \frac{h^3}{3!} y_2^{III} + \frac{h^4}{4!} y_2^{IV} + \ldots\ldots \qquad \rightarrow (4)$$

In general, Taylor's expansion of y(x) at a point $x = x_n$ is

$$y_{n+1} = y_n + \frac{h}{1!} y_n^{I} + \frac{h^2}{2!} y_n^{II} + \frac{h^3}{3!} y_n^{III} + \frac{h^4}{4!} y_n^{IV} + \ldots\ldots \qquad \rightarrow (5)$$

Example1. Using Taylor's expansion evaluate the integral of $y' - 2y = 3e^x, y(0) = 0$ at $x = 0.2$. Hence compare the numerical solution obtained with exact solution.

Sol: Given equation can be written as $2y + 3e^x = y', y(0) = 0$

Differentiating repeatedly w.r.t to 'x' and evaluating at $x = 0$

$$y'(x) = 2y + 3e^x, y'(0) = 2y(0) + 3e^0 = 2(0) + 3(1) = 3$$
$$y''(x) = 2y' + 3e^x, y''(0) = 2y'(0) + 3e^0 = 2(3) + 3 = 9$$
$$y'''(x) = 2.y''(x) + 3e^x, y'''(0) = 2y''(0) + 3e^0 = 2(9) + 3 = 21$$
$$y^{iv}(x) = 2.y'''(x) + 3e^x, y^{iv}(0) = 2(21) + 3e^0 = 45$$
$$y^{v}(x) = 2.y^{iv} + 3e^x, y^{v}(0) = 2(45) + 3e^0 = 90 + 3 = 93$$

In general, $y^{(n+1)}(x) = 2.y^{(n)}(x) + 3e^x$ or $y^{(n+1)}(0) = 2.y^{(n)}(0) + 3e^0$

The Taylor's series expansion of $y(x)$ about $x_0 = 0$ is

$$y(x) = y(0) + xy'(0) + \frac{x^2}{2!} y''(0) + \frac{x^3}{3!} y'''(0) + \frac{x^4}{4!} y''''(0) + \frac{x^5}{5!} y'''''(0) + \ldots.$$

Substituting the values of $y(0), y'(0), y''(0), y'''(0), \ldots\ldots\ldots.$

$$y(x) = 0 + 3x + \frac{9}{2}x^2 + \frac{21}{6}x^3 + \frac{45}{24}x^4 + \frac{93}{120}x^5 + \ldots\ldots..$$

$$y(x) = 3x + \frac{9}{2}x^2 + \frac{7}{2}x^3 + \frac{15}{8}x^4 + \frac{31}{40}x^5 + \ldots\ldots.. \rightarrow \text{equ1}$$

Now put $x = 0.1$ in equ1

$$y(0.1) = 3(0.1) + \frac{9}{2}(0.1)^2 + \frac{7}{2}(0.1)^3 + \frac{15}{8}(0.1)^4 + \frac{31}{40}(0.1)^5 = 0.34869$$

Now put $x = 0.2$ in equ1

$$y(0.2) = 3(0.2) + \frac{9}{2}(0.2)^2 + \frac{7}{2}(0.2)^3 + \frac{15}{8}(0.2)^4 + \frac{31}{40}(0.2)^5 = 0.811244$$

$$y(0.3) = 3(0.3) + \frac{9}{2}(0.3)^2 + \frac{7}{2}(0.3)^3 + \frac{15}{8}(0.3)^4 + \frac{31}{40}(0.3)^5 = 1.41657075$$

Analytical Solution:

The exact solution of the equation $\frac{dy}{dx} = 2y + 3e^x$ with $y(0) = 0$ can be found as follows

$\frac{dy}{dx} - 2y = 3e^x$ Which is a linear in y.

Here $P = -2, Q = 3e^x$

I.F $= \int_e^{pdx} = \int_e^{-2dx} = e^{-24}$

General solution is $y.e^{-2x} = \int 3e^x.e^{-2x}dx + c = -3e^{-x} + c$

$\therefore y = -3e^x + ce^{2x}$ where $x = 0, y = 0 \quad 0 = -3 + c \Rightarrow c = 3$

The particular solution is $y = 3e^{2x} - 3e^x$ or $y(x) = 3e^{2x} - 3e^x$

Put $x = 0.1$ in the above particular solution,

$y = 3.e^{0.2} - 3e^{0.1} = 0.34869$

Similarly put $x = 0.2$, $y = 3e^{0.4} - 3e^{0.2} = 0.811265$

put $x = 0.3$, $y = 3e^{0.6} - 3e^{0.3} = 1.416577$

EULER'S METHOD

It is the simplest one-step method and it is less accurate. Hence it has a limited application.

Consider the differential equation $\frac{dy}{dx}$ = f(x,y) $\rightarrow$(1)

With $y(x_0) = y_0$ $\rightarrow$(2)

Consider the first two terms of the Taylor's expansion of y(x) at $x = x_0$

$y(x) = y(x_0) + (x - x_0)\, y^1(x_0)$ $\rightarrow$(3)

from equation (1) $y^1(x_0) = f(x_0,y(x_0)) = f(x_0,y_0)$

Substituting in equation (3)

$\therefore y(x) = y(x_0) + (x - x_0)\, f(x_0,y_0)$

At $x = x_1$, $y(x_1) = y(x_0) + (x_1 - x_0)\, f(x_0,y_0)$

$\therefore y_1 = y_0 + h\, f(x_0,y_0)$ where $h = x_1 - x_0$

Similarly at $x = x_2$, $y_2 = y_1 + h\, f(x_1,y_1)$,

Proceeding as above, $\mathbf{y_{n+1} = y_n + h\, f(x_n,y_n)}$

This is known as Euler's Method

Example 1. Using Euler's method solve for x = 2 from $\frac{dy}{dx} = 3x^2 + 1, y(1) = 2$, taking step size (I) h = 0.5 and (II) h=0.25

Sol: Here $f(x,y) = 3x^2 + 1$, $x_0 = 1, y_0 = 2$

Euler's algorithm is $y_{n+1} = y_n + h\, f(x_n,y_n)$, n = 0,1,2,3,..... →(1)

h = 0.5 $\therefore x_1 = x_0 + h = 1+0.5 = 1.5$

Taking n = 0 in (1), we have $x_2 = x_1 + h = 1.5 + 0.5 = 2$

$y_1 = y_0 + h\, f(x_0,y_0)$

i.e. $y_1 = y(0.5) = 2 + (0.5)\, f(1,2) = 2 + (0.5)\,(3 + 1) = 2 + (0.5)(4)$

Here $x_1 = x_0 + h = 1 + 0.5 = 1.5$

$\therefore y(1.5) = 4 = y_1$

Taking n = 1 in (1),we have

$y_2 = y_1 + h\, f(x_1,y_1)$

i.e. $y(x_2) = y_2 = 4 + (0.5)\, f(1.5,4) = 4 + (0.5)[3(1.5)^2 + 1] = 7.875$

Here $x_2 = x_4 + h = 1.5 + 0.5 = 2$

$\therefore y(2) = 7.875$

h = 0.25 $\therefore x_1 = 1.25,\ x_2 = 1.50,\ x_3 = 1.75,\ x_4 = 2$

Taking n = 0 in (1), we have

$y_1 = y_0 + h\, f(x_0,y_0)$

i.e. $y(x_1) = y_1 = 2 + (0.25)\, f(1,2) = 2 + (0.25)\,(3 + 1) = 3$

$y(x_2) = y_2 = y_1 + h\, f(x_1,y_1)$

i.e. $y(x_2) = y_2 = 3 + (0.25)\, f(1.25,3)$

$= 3 + (0.25)[3(1.25)^2 + 1]$

$= 4.42188$

Here $x_2 = x_1 + h = 1.25 + 0.25 = 1.5$

$\therefore y(1.5) = 5.42188$

Taking n = 2 in (1), we have

i.e. $y(x_3) = y_3 = h\ f(x_2,y_2)$

$= 5.42188 + (0.25)\ f(1.5,2)$

$= 5.42188 + (0.25)\ [3(1.5)^2 + 1]$

$= 6.35938$

Here $x_3 = x_2 + h = 1.5 + 0.25 = 1.75$

$\therefore y(1.75) = 7.\ 35938$

Taking n = 4 in (1),we have

$y(x_4) = y_4 = y_3 + h\ f(x_3,y_3)$

i.e. $y(x_4) = y_4 = 7.35938 + (0.25)\ f(1.75,2)$

$= 7.35938 + (0.25)[3(1.75)^2 + 1]$

$= 8.90626$

Note that the difference in values of y(2) in both cases

(i.e. when h = 0.5 and when h = 0.25).The accuracy is improved significantly when h is reduced to 0.25 (Exact solution of the equation is $y = x^3 + x$ and with this $y(2) = y_2 = 10$.

Modified Euler's method

It is given by $y^{(i)}_{k+1} = y_k + h/2 f\left[\left(x_k, y_k\right) + f\left(x_{k+1}, 1\right)_{k+1}{}^{(i-1)}\right], i = 1, 2....., ki = 0, 1.....$

Working rule :

i) Modified Euler's method

$y^{(i)}_{k+1} = y_k + h/2 f\left[\left(x_k, y_k\right) + f\left(x_{k+1}, 1\right)_{k+1}{}^{(i-1)}\right], i = 1, 2....., ki = 0, 1.....$

ii) When $i = 1$ y^{0}_{k+1} can be calculated from Euler's method

iii) k=0, 1……… gives number of iteration. $i = 1, 2...$

gives number of times, a particular iteration k is repeated

Suppose consider dy/dx=f(x, y) -------- (1) with $y(x_0) = y_0$----------- (2)2

To find $y(x_1) = y_1$ at $x = x_1 = x_0 + h$

Now take k=0 in modified Euler's method

We get $y_1^{(1)} = y_0 + h/2\left[f\left(x_0, y_0\right) + f\left(x_1, y_1^{(i-1)}\right)\right]$……………………… (3)

Taking i=1, 2, 3...k+1 in equation (3), we get

$y_1^{(0)} = y_0 + h/2\left[f\left(x_0, y_0\right)\right]$ (By Euler's method)

$y_1^{(1)} = y_0 + h/2\left[f\left(x_0, y_0\right) + f\left(x_1, y_1^{(0)}\right)\right]$

$y_1^{(2)} = y_0 + h/2\left[f\left(x_0, y_0\right) + f\left(x_1, y_1^{(1)}\right)\right]$

$y_1^{(k+1)} = y_0 + h/2\left[f\left(x_0, y_0\right) + f\left(x_1, y_1^{(k)}\right)\right]$

If two successive values of $y_1^{(k)}, y_1^{(k+1)}$ are sufficiently close to one another, we will take the common value as $y_2 = y\left(x_2\right) = y\left(x_1 + h\right)$

We use the above procedure again

Example1. Using modified Euler's method, find the approximate value of x when $x = 0.3$ given that $dy/dx = x + y \text{ and } y\left(0\right) = 1$

Sol: Given $dy/dx = x + y \text{ and } y\left(0\right) = 1$

Here $f\left(x, y\right) = x + y, x_0 = 0, \text{ and } y_0 = 1$

Take h = 0.1 which is sufficiently small

Here $x_0 = 0, x_1 = x_0 + h = 0.1, x_2 = x_1 + h = 0.2, x_3 = x_2 + h = 0.3$

The formula for modified Euler's method is given by

$$y_{k+1}^{(i)} = y_k + h/2\left[f\left(x_k + y_k\right) + f\left(x_{k+1}, y_{k+1}^{(i-1)}\right)\right] \rightarrow (1)$$

<u>Step1:</u> To find $y_1 = y(x_1) = y\ (0.1)$

Taking k = 0 in eqn(1)

$$y_{k+1}^{(i)} = y_0 + h/2\left[f\left(x_0 + y_0\right) + f\left(x_1, y_1^{(i-1)}\right)\right] \rightarrow (2)$$

when i = 1 in eqn (2)

$$y_1^{(i)} = y_0 + h/2\left[f\left(x_0, y_0\right) + f\left(x_1, y_1^{(0)}\right)\right]$$

First apply Euler's method to calculate $y_1^{(0)}$ = y_1

$\therefore \quad y_1^{(0)} = y_0 + h\, f\left(x_0, y_0\right)$

= 1+(0.1)f(0.1)

= 1+(0.1)

= 1.10

$$now\left[x_0 = 0, y_0 = 1, x_1 = 0.1, y_1(0) = 1.10\right]$$

$$\therefore\ y_1^{(1)} = y_0 + 0.1/2\left[f\left(x_0, y_0\right) + f\left(x_1, y_1^{(0)}\right)\right]$$

= 1+0.1/2[f(0,1) + f(0.1,1.10)

= 1+0.1/2[(0+1)+(0.1+1.10)]

= 1.11

When i=2 in equation (2)

$$y_1^{(2)} = y_0 + h/2\left[f\left(x_0, y_0\right) + f\left(x_1, y_1^{(1)}\right)\right]$$

= 1+0.1/2[f(0.1)+f(0.1,1.11)]

= 1 + 0.1/2[(0+1)+(0.1+1.11)]

= 1.1105

$$y_1^{(3)} = y_0 + h/2\left[f\left(x_0, y_0\right) + f\left(x_1, y_1^{(2)}\right)\right]$$

= 1+0.1/2[f(0,1)+f(0.1 , 1.1105)]

= 1+0.1/2[(0+1)+(0.1+1.1105)]

= 1.1105

Since $y_1^{(2)} = y_1^{(3)}$

$\therefore$ y_1 = 1.1105

Step:2 To find $y_2 = y(x_2) = y(0.2)$

Taking k = 1 in equation (1) , we get

$$y_2^{(i)} = y_1 + h/2\left[f\left(x_1, y_1\right) + f\left(x_2, y_2^{(i-1)}\right)\right] \rightarrow (3)$$

I = 1,2,3,4,.....

For i = 1

$$y_2^{(1)} = y_1 + h/2\left[f\left(x_1, y_1\right) + f\left(x_2, y_2^{(0)}\right)\right]$$

$y_2^{(0)}$ is to be calculate from Euler's method

$$y_2^{(0)} = y_1 + h\ f\left(x_1, y_1\right)$$

= 1.1105 + (0.1) f(0.1 , 1.1105)

= 1.1105+(0.1)[0.1+1.1105]

= 1.2316

$$\therefore\ y_2^{(1)} = 1.1105 + 0.1/2\left[f\left(0.1, 1.1105\right) + f\left(0.2, 1.2316\right)\right]$$

= 1.1105 +0.1/2[0.1+1.1105+0.2+1.2316]

= 1.2426

$$y_2^{(2)} = y_1 + h/2\left[f\left(x_1, y_1\right) + f\left(x_2 y_2^{(1)}\right)\right]$$

= 1.1105 + 0.1/2[f(0.1 , 1.1105) , f(0.2 . 1.2426)]

= 1.1105 + 0.1/2[1.2105 + 1.4426]

= 1.1105 + 0.1(1.3266)

= 1.2432

$$y_2^{(3)} = y_1 + h/2\left[f\left(x_1, y_1\right) + f\left(x_2 y_2^{(2)}\right)\right]$$

= 1.1105+0.1/2[f(0.1,1.1105)+f(0.2 , 1.2432)]

= 1.1105+0.1/2[1.2105+1.4432)]

= 1.1105 + 0.1(1.3268)

= 1.2432

Since $y_2^{(3)} = y_2^{(3)}$

Hence y_2 = 1.2432

Step:3

To find $y_3 = y(x_3)$ = y y(0.3)

Taking k =2 in equation (1) we get

$$y_3^{(1)} = y_2 + h/2\left[f\left(x_2, y_2\right) + f\left(x_3, y_3^{(i-1)}\right)\right] \rightarrow (4)$$

For i = 1 ,

$$y_3^{(1)} = y_2 + h/2\left[f\left(x_2, y_2\right) + f\left(x_3, y_3^{(0)}\right)\right]$$

$y_3^{(0)}$ is to be evaluated from Euler's method .

$$y_3^{(0)} = y_2 + h\, f\left(x_2, y_2\right)$$

= 1.2432 +(0.1) f(0.2 , 1.2432)

= 1.2432+(0.1)(1.4432)

= 1.3875

$\therefore\ y_3^{(1)}$ = 1.2432+0.1/2[f(0.2 , 1.2432)+f(0.3, 1.3875)]

= 1.2432 + 0.1/2[1.4432+1.6875]

= 1.2432+0.1(1.5654)

= 1.3997

$$y_3^{(2)} = y_2 + h/2\left[\ f\left(x_2, y_2\right) + f\left(x_3, y_3^{(1)}\right)\right]$$

= 1.2432+0.1/2[1.4432+(0.3+1.3997)]

= 1.2432+ (0.1) (1.575)

= 1.4003

$$y_3^{(3)} = y_2 + h/2\left[\ f\left(x_2, y_2\right) + f\left(x_3, y_3^{(2)}\right)\right]$$

= 1.2432+0.1/2[f(0.2 , 1.2432)+f(0.3 , 1.4003)]

= 1.2432 + 0.1(1.5718)

= 1.4004

$$y_3^{(4)} = y_2 + h/2\left[\ f\left(x_2, y_2\right) + f\left(x_3, y_3^{(3)}\right)\right]$$

= 1.2432 + 0.1/2[1.4432+1.7004]

= 1.2432+(0.1)(1.5718)

= 1.4004

Since $y_3^{(3)} = y_3^{(4)}$

$\therefore$ The value of y at x = 0.3 is 1.4004

Runge – Kutta Methods

I. Second order R-K Formula

$y_{i+1} = y_i + 1/2\ (K_1 + K_2)$,

Where $K_1 = h\ (x_i, y_i)$

$K_2 = h\ (x_i + h,\ y_i + k_1)$ for i= 0,1,2-------

II. Third order R-K Formula

$y_{i+1} = y_i+1/6\ (K_1+4K_2+ K_3)$,

Where $K_1 = h\ (x_i, y_i)$

$K_2 = h\ (x_i+h/2, y_0+k_1/2)$

$K_3 = h\ (x_i+h, y_i+2k_2-k_1)$

For i= 0,1,2-------

III. Fourth order R-K Formula

$y_{i+1} = y_i+1/6\ (K_1+2K_2+ 2K_3 +K_4)$,

Where $K_1 = h\ (x_i, y_i)$

$K_2 = h\ (x_i+h/2, y_i+k_1/2)$

$K_3 = h\ (x_i+h/2, y_i+k_2/2)$

$K_4 = h\ (x_i+h, y_i+k_3)$

For i= 0,1,2-------

Example 1. Apply the 4^{th} order R-K method to find an approximate value of y when x=1.2 in stepsof 0.1, given that $y^1 = x^2+y^2$, y (1)=1.5

sol. Given $y^1= x^2+y^2$,and y(1)=1.5

Here $f(x,y)= x^2+y^2$, $y_0 =1.5$ and $x_0=1$,h=0.1

So that $x_1=1.1$ and $x_2=1.2$

Step1:

To find y_1 :

By 4^{th} order R-K method we have

$y_1=y_0+1/6\ (k_1+2k_2+2k_3+k_4)$

$k_1= h\ f(x_0,y_0) = (0.1)\ f(1,1.5) = (0.1)\ [1^2+(1.5)^2] = 0.325$

$k_2= hf\ (x_0+h/2,y_0+k_1/2) = (0.1)\ f(1+0.05,1.5+0.325) = 0.3866$

and

$k_3= h\ f((x_0+h/2,y_0+k_2/2) = (0.1)\ f(1.05,1.5+0.\ 3866/2) = (0.1)[(1.05)^2+(1.6933)^2]$

$= 0.39698$

$k_4= hf(x_0+h,y_0+k_3) = (0.1)f(1.0,1.89698) = 0.48085$

Hence

$$y_1 = 1.5 + \frac{1}{6}\left[0.325 + 2(0.3866) + 2(0.39698) + 0.48085\right]$$
$$= 1.8955$$

Step2:

To find y_2, i.e., $y(x_2) = y(1.2)$

Here x_1=0.1,y_1=1.8955 and h=0.1

by 4^{th} order R-K method we have

$y_2 = y_1+(1/6)(k_1+2k_2+2k_3+k_4)$

$k_1= hf(x_1,y_1) = (0.1)f(0.1,1.8955) = (0.1)[1^2+(1.8955)^2] = 0.48029$

$k_2 = hf(x_1+h/2,y_1+k_1/2) = (0.1)f(1.1+0.1,1.8937+0.4796) = 0.58834$

$k_3 = hf((x_1+h/2,y_1+k_2/2) = (0.1)f(1.5,1.8937+0.58743) = (0.1)[(1.05)^2+(1.6933)^2]$

$= 0.611715$

$k_4 = hf(x_1+h,y_1+k_3) = (0.1)f(1.2,1.8937+0.610728) = 0.77261$

Hence

$y_2 = 1.8937+(1/6)(0.4796+2(0.58834)+2(0.611715)+0.7726) = 2.5043$

$\therefore$ y =2.5043 where $x = 0.2$.

NUMERICAL AND STATISTICAL METHODS

UNIT IV

PROBABILITY AND EXPECTATION OF RANDOM VARIABLE

Objectives:

(a) To understand the concepts of probability and statistics.
(b) To know sampling theory and principles of hypothesis testing
(c) To appreciate Queuing theory and models.

Syllabus :

(a) Axioms of probability (Non-negativity, Totality, and Additivity)
(b) Conditional and Unconditional probabilities (Definitions and simple problems)
(c) Additive law of probability (simple applications)
(d) Multiplicative law of probability (simple applications)
(e) Baye's Theorem (without proof and applications)
(f) Concept of a Random variable (one dimensional case definition only and simple examples)
(g) Types of random variables (Discrete and Continuous cases)
(h) Probability mass function and probability density function – their properties (without proofs)
(i) Distribution Function and its properties (without proofs)
(j) Evaluation of mean and variance (problems)

Learning Outcomes:

Students will be able to

(a) understand the usage of axioms of probability
(b) apply various laws of probability (like additive, multiplicative, and Baye's) in real-life problems
(c) distinguish between discrete random variable (DRV) and continuous random variable (CRV)
(d) understand the complexity in finding the mean and the variance of DRV and CRV.

Learning Material

Probability and Expectation of Random Variable

Terminology associated with Probability Theory:

Random experiment: If an experiment or trial can be repeated any number of times under similar conditions and it is possible to enumerate the total number of outcomes, but an individual outcome is not predictable, such an experiment is called a random experiment. For instance, if a fair coin is tossed three times, it is possible to enumerate all the possible eight sequences of head (H) and tail (T). But it is not possible to predict which sequence will occur at any occasion.

***Outcome*:** A result of an experiment is termed as an outcome. Head (H) and tail (T) are the outcomes when a fair coin is flipped.

Sample Space: Each conceivable outcome of a random experiment under consideration is said to be a sample point. The totality of all conceivable sample points is called a sample space. In other words, the list of all possible outcomes of an experiment is called a sample space. For example, the set $\{HH, HT, TH, TT\}$ constitutes a sample space when two fair coins tossed at a time.

(i) *Discrete Sample Space:* A sample space which consists of countably finite or infinite elements or sample points is called discrete sample space. It is abbreviated as DSS.

(ii) *Continuous Sample Space:* A sample space which consists of continuum of values is called continuous sample space. It is abbreviated as CSS.

***Event*:** Any subset of the sample space is an event. In other words, the set of sample points which satisfy certain requirement(s) is called an event. For example, in the event, there are exactly two heads in three tossing's of a coin, it would consist of three points (H, H, T), (H, T, H), and (T, H, H). Each point is called an event. i.e. an outcome which further cannot be divided is called an event. Events are classified as:

***Elementary Event*:** An event or a set consists only one element or sample point is called an elementary event. It is also termed as simple event.

***Complementary Event*:** Let A be the event of S. The non-occurrence of A and contains those points of the sample space which do not belong to A.

***Exhaustive Events*:** All possible events in any trial are known as exhaustive events. In tossing a coin, there are two exhaustive elementary events namely, head and tail.

Equally Likely Events: Events are said to be equally when there is no reason to expect anyone of them rather than anyone of the others in a single trial of the random experiment. In other words, all the sample units or outcomes of sample space are having equal preference to each other, then the events are said to be equally likely events. In a tossing a coin, the outcomes head (H) and tail (T) are equally likely events.

Mutually Exclusive Events: Events A and B are said to be mutually exclusive events if the occurrence of A precludes the occurrence of B and vice-versa. In other words, if there is no sample point in A which is common to the sample point in B, i.e. $A \cap B = \phi$, the events A and B are said to be mutually exclusive. For example, if we flip a fair coin, we find either H or T in a trial, but not both. i.e. happening of H that prevents the happening of T in a trial, then H and T are mutually exclusive events. (No two events can happen simultaneously in a trial, such events are mutually exclusive.)

Independent events: Two events A and B are said to be independent if the occurrence of A has no bearing on the occurrence of B i.e. the knowledge that the event A has occurred gives no information about the occurrence of the event B.

Formally, two events A and B are independent if and only if,

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

For example, a bag contains balls of two different colours say, red and white. The two balls are drawn successively. First a ball is drawn from one bag and replaced after noting its color. Let us presume that it is white and is denoted by the event A. Another ball is drawn from the same bag and its colour is noted. Let this event noted by the event B. The result of the second drawn is not influenced by the first drawn. Hence the events A and B are said to be independent.

Various definitions of probability are 1. Classical definition of probability (or Mathematical definition of probability) 2. Statistical definition of probability (or Empirical definition of probability) 3. Axiomatic approach to probability.

The classical definition of probability breaks down when we do not have a complete priori analysis i.e. when the outcomes of the trial are not equally or when the total number of trials is infinite or when the enumeration of all equally likely events is not possible. So the necessity of the statistical definition of probability arises.

The statistical definition of probability, although is of great use from practical point of view, is not conductive for mathematical approach since an actual limiting number may not really exist. Hence another definition is thought of based on axiomatic approach. This definition leads to the development of calculus of probability.

- P(E) = Favourable number of cases / Total cases

- Limits of probability $0 \le$ P (E) ≤ 1

Axiomatic approach to Probability: A real valued function $p(x): S(x) \to (0,1)$ is called a probability function which satisfies the following rules or statements technically termed as axioms, where S is a sample space, x is a result of an experiment ranges from $-\infty$ to ∞.
Axiom 1: (Non-negativity) For any event E of S, P(E) ≥ 0.
Axiom 2: (Totality) S be a sample space of an experiment and P(S) = 1.
Axiom 3: (Additive) Suppose E_1 and E_2 be mutually exclusive events of S, then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$.

For example, police department needs new tires for its patrol cars and the probabilities are 0.17, 0.22, 0.03, 0.29, 0.21, and 0.08 that it will buy Uniroyal tires, Goodyear tires, Michelin tires, General tires, Goodrich tires or Armstrong tires. Then the probabilities that the combinations of (Goodyear, Goodrich), (Uniroyal, General, Goodrich), (Michelin, Armstrong), and (Goodyear, General, Armstrong) tires respectively are 0.43, 0.67, 0.11, and 0.59 respectively.

Note: Axioms of probability do not determine probabilities. But the axioms restrict the assignments of probabilities in a manner that enables us to interpret probabilities as relative frequencies without inconsistencies.

Unconditional Probability:

The individual probabilities of the events of A and B are termed as unconditional probabilities. i.e. unconditional probabilities are the probabilities which are not influenced by other events in the sample space, S. These are also termed as priori probabilities.

Result : If A is any event in s, then $P(\overline{A}) = 1 - P(A)$

Result : For any two events A and B then

(i) $P(\overline{A} \cap B) = P(B) - P(A \cap B)$

(ii) $P(A \cap \overline{B}) = P(A) - P(A \cap B)$

Additive Law of Probability:

Statement: If A and B are any two events of a sample space S and are not disjoint then

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

Proof:

$$A \cup B = A \cup (\overline{A} \cap B)$$

$$\begin{aligned} P(A \cup B) &= P[A \cup (\overline{A} \cap B)] \\ &= P(A) + P(\overline{A} \cap B) \\ &= P(A) + [P(\overline{A} \cap B) + P(A \cap B) - P(A \cap B)] \\ &= P(A) + P[(\overline{A} \cap B) \cup (A \cap B)] - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

Result : If A and B are disjoint events then $P(A \cup B) = P(A) + P(B)$

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

Solution: We know that (WKT), a pack of playing cards consists 52 in number. i.e. the sample space of pack of cards, n(S) = 52.

Let A denoted the event of getting a spade and B denotes the event of getting an ace. Then the probability of the event of getting either a spade or an ace is

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

Since all the cards are equally likely, mutually exclusive, we have

$$P(A) = \frac{13}{52},\ P(B) = \frac{4}{52},\ P(A \cap B) = \frac{1}{52}$$

By addition theorem of probability,

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

Conditional Probability: In many situations arises in our day to day life about the occurrence of an event A (for instance, getting treatment) is influenced by the occurrence of the event B (availability of doctor) and the event is known a conditional event, denoted by $A|B$ and hence the probability of the conditional event is known as 'conditional probability' and is denoted by $P(A|B)$.

Definition: Let A and B be two events. The conditional probability of event B, if an event A has occurred, is defined by the relation,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)},\ if\ P(A) > 0.$$

i.e. the conditional probability of the event B is the ratio of the probability of the joint occurrence of the events A and B to the unconditional probability of the event A.

Similarly, we can define $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)},\ if\ P(B) > 0$.

Example: In a group consisting of men and women are equal in number. 10% of the men and 45% of the women are unemployed. If a person is selected randomly from the group then find the probability that the person is an unemployed.

Solution: Since the men and women are equal in number in a group, we take

$$P(M) = 1/2 \text{ and } P(W) = 1/2$$

Let E be the event of employed person. Then $\bar{E}$ be the event of unemployed.

Then we have, $P(E \mid M) = 10\% = 0.10$, $P(E \mid W) = 40\% = 0.45$

implies, $P(\bar{E} \mid M) = 0.90$, $P(\bar{E} \mid W) = 0.55$

The probability that the person (either male or female) is an unemployed is

$$P(\bar{E}) = P(M)\,P(\bar{E} \mid M) + P(W)\,P(\bar{E} \mid W)$$

$$= \frac{1}{2}(0.90) + \frac{1}{2}(0.55) = 0.725$$

Example: Two marbles are drawn in succession from a box containing 10 red, 30 white, 20 blue and 15 orange marbles with replacement being made after each draw. Find the probability that (i) both are white (ii) first is red and second is white.

Solution: From the given information, we noticed that

the number of marbles in the box = 75.

i. Let us define E_1 be the event of 1st drawn marble is white and E_2 be the event of 2nd drawn marble is also white.

   Since we are using with replacement to select marbles in succession, we have

   $$P(E_1) = \frac{30}{75} \text{ and } P(E_2) = \frac{30}{75}$$

   Therefore, the probability that both marbles are white is

   $$P(E_1 \cap E_2) = P(E_1)\,P(E_2 \mid E_1) = \frac{30}{75}.\frac{30}{75} = \frac{4}{25}$$

ii. Let us define E_1 be the event of 1st drawn marble is red and E_2 be the event of 2nd drawn marble is white.

   Since we are using with replacement to select marbles in succession, we have

   $$P(E_1) = \frac{10}{75} = \frac{2}{15} \text{ and } P(E_2) = \frac{30}{75} = \frac{2}{5}$$

   Therefore, the probability that the first draw marble is red and the second draw marble is white is

   $$P(E_1 \cap E_2) = P(E_1)\,P(E_2 \mid E_1) = \frac{2}{15}.\frac{2}{5} = \frac{4}{75}.$$

Multiplicative Law of Probability:

Statement: For any events A and B in the sample space S, we have

$$P(A \cap B) = P(A)\,P(B\,|\,A),\ P(A) > 0$$
$$= P(B)\,P(A\,|\,B)\ ,\ P(B) > 0$$

Where $P(B\,|\,A)$ is the conditional probability of B provided A has already happened and $P(A\,|\,B)$ is the conditional probability of A provided B has already happened.

Result: If A and B are independent events then P(A ∩ B)= P (A) P(B)

Result: If A1,A2……..An are n independent events then probability of happening of at least one of the event= 1- probability of none of the events happening

Baye's Theorem:

Statement: Suppose $E_1, E_2, ..., E_n$ be 'n' mutually exclusive events in S with $P(E_i) \neq 0;\ i = 1, 2, ..., n.$ Let A be any arbitrary event which is a subset of S and $P(A) > 0.$

Then, we have $P(E_i\,|\,A) = \dfrac{P(E_i)\,P(A\,|\,E_i)}{\sum_{i=1}^{n} P(E_i)\,P(A\,|\,E_i)}$, $i = 1, 2, ..., n$.

Where $P(E_i)'s$ are called 'a priori probabilities', $P(A\,|\,E_i)'s$ are called 'likelihoods' and $P(E_i\,|\,A)'s$ are called 'posterior probabilities'.

Note: $\sum_{i=1}^{n} P(E_i)\,P(A\,|\,E_i)$ = P(A) is called Total probability

Example: Four computer companies A, B, C and supply transistors to a company. From previous experience, it is known that the probability of the transistors being bad if it comes from A is 40%, from B is 2%, from C is 5% and from D is 1%. The probabilities of picking supplier A is 20%, B is 30%, C is 10% and D is 40%.

(i) Find the probability that a transistor chosen at random is bad.

(ii) Find the probability that the transistor comes from company A, given that the transistor is bad.

Sol: Probabilities of picking suppliers A,B,C and D are $P(E_1) = 0.2$, $P(E_2) = 0.3$, $P(E_3) = 0.1$, and $P(E_4) = 0.4$ respectively

Getting *suppliers, A,B,C and D when picked are events E1,E2,E3 and E4 respectively*

D is the event bad

Given $P(D/E_1) = 0.4$, $PD/(E_2) = 0.02$, $P(D/E_3) = 0.05$, $P(D/E_4) = 0.01$

(i) $P(D)= P(E_1)\,P(D/E_1) + P(E_2)\,P(D/E_2) + P(E_3)\,P(D/E_3) + P(E_4)\,P(D/E_4)$

$= 0.895$

(ii) $P(E_1/A)= P(E_1)\,P(D/E_1) / P(D) = 0.893$

Concept of Random Variable:

A random variable X is a real function of the events of a given sample space S. Thus for a given experiment defined by a sample space S with events s, the random variable is a function of s. It is denoted by X(s). A random variable X can be considered to be a function that maps all events of the sample space into points on the real axis.

For example, an experiment consists of tossing two coins. Let the random variable be a function X chosen as the number of heads shown. So X maps the real numbers of the event "showing no head" as zero, the event "any one is head" as one and "both heads" as two. Therefore, the random variable is X = $\{0,1,2\}$. The elements of the random variable X are $x_1 = 0,\ x_2 = 1,\ and\ x_3 = 2$.

Types of Random Variable:

Random variables are classified into:

1. *Discrete Random Variable* (DRV): A random variable X which is defined on the discrete sample space is called discrete random variable.
   For example, consider a discrete sample space $S = \{1,2,3,4\}$. Let us define X = S^2 be a random variable. Then discrete values of S map to discrete values of X as $\{1,4,9,16\}$. The probabilities of the random variable x are equal to the probabilities of set S because of the one-to-one mapping of the discrete points.
   Let X be a discrete random variable with integer events $X = \{x_1, x_2, ..., x_n\}$. The probability of X at any event is a function of x_i and is given by
   $$P(X = x_i) = p(x_i)\ ,\ i = 1,2,3,.......$$
   This function is called probability mass function and is abbreviated as p.m.f. (pmf).

Properties of probability mass function:

Consider a discrete random variable X in a sample space with infinite number of possible outcomes, that is, $X = \{x_1, x_2, ...\}$. If the probability of X, $p(x_i)$, $i = 1,2,3,....$ satisfies the following properties then the function $p(x)$ is called probability mass function.

(i) $p(x_i) \geq 0,\ \forall i$ (ii) $\sum_{i=1}^{\infty} p(x_i) = 1$

2. *Continuous Random Variable* (CRV): A random variable X which is defined on the continuous sample space is called continuous random variable.
   Temperature, time, height and weight over a period of time etc. are examples of CRV.
   The probability density function of a random variable X is defined as the variable X falls in the infinitesimal interval $\left[x - \frac{dx}{2}, x + \frac{dx}{2}\right]$ such that $P\left(x - \frac{dx}{2} \leq X \leq x + \frac{dx}{2}\right) = f(x)dx$,

$$i.e.\ f(x) = \underset{\Delta t \to 0}{Lt} \frac{P\left(x - \frac{dx}{2} \le X \le x + \frac{dx}{2}\right)}{dx}$$

Where $f(x)$ is called the probability density function of a random variable X and the continuous curve y = f(x) is called probability density curve.

Properties of probability density function:

The continuous curve y =f(x) satisfies the following properties, then the function f(x) is called probability density function of a random variable and is abbreviated as p.d.f. (pdf).

(i) $f(x) \ge 0,\ \forall\, x,$ (ii) $\int_{-\infty}^{\infty} f(x)dx = 1$

3. *Mixed Random Variable*: A random variable which is defined on both DSS and CSS partially, then the random variable is said to be a mixed random variable.

Probability distribution function:

Let X be a random variable. Then the probability distribution function associated with X is defined as the probability that the outcomes of an experiment will be one of the outcomes for which $X(s) \le x, x \in R.$ That is, the function $F(x)$ is defined by

$F(x) = P(X \le x) = P\{s : X(s) \le x\}, -\infty < x < \infty$ is called the distribution function of X.

Sometimes it is also known as Cumulative Distribution function and is abbreviated as CDF.

Properties of cdf:

1 . If F is the distribution function of a random variable S and $a < b,$ then

(i) $P(a < X \le b) = F(b) - F(a)$

(ii) $P(a \le X \le b) = P(X = a) + [F(b) - F(a)]$

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

(iv) $P(a \le X \le b) = [F(b) - F(a)] = P(X = b) + P(X = a)$

2. All distribution functions are monotonically increasing and lie between 0 and 1. That is, if F is the distribution function of the random variable X, then

(i) $0 \le F(x) \le 1$ i.e. F is bounded.

(ii) $F(x) < F(y)$ when $x < y.$

(ii) $F(-\infty) = 0$ and $F(\infty) = 0.$

Evaluation of mean and variance:

1. A random variable ‘X’ has the following probability functions:

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

(i) Determine ‘K’ (ii) Evaluate $P(X < 6), (PX \ge 6) \& Plo < x < 5$

(iii) Mean (iv) Variance

Sol: $Since \sum_{x=0}^{7} P(x) = 1$

K = 1/10 = 0.1

P(X<6)=P(X=0)+P(x=1)+……………..+P(x=5)

=0.81

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

$Mean\,\mu = \sum_{i=0}^{7} P_i x_i$

= 3.66 (K=1/10)

Variance $= \sum_{i=0}^{7} P_i x_i^{\ 2} - \mu^2$

=3.4044

2. The probability density f(x) of a continuous random variable is given by f(x) = C.e$^{-|x|}$, $-\infty < x < \infty$. S.T C = ½ and find (i) mean (ii) variance of the distribution. Also find $P(0 \le X \le 4)$.

Sol: We have $\int_{-\infty}^{\infty} f(x)dx = 1$

$\Rightarrow C = 1/2$

(i) Mean $\mu = \int_{-\infty}^{\infty} xf(x)dx. = \frac{1}{2}\int_{-\infty}^{\infty} x.e^{-|x|}dx = \infty$

= 0 [integrand is odd]

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

=2

(iii) $P(0 \le X \le 4)$=0.49

$\int_{0}^{4} f(x)dx = 0.49$

UNIT-5
Probability distribution and Correlation & Regression

Objectives:

- To know the importance of correlation coefficient & lines of regression.

Syllabus:

Review on Binomial and Poisson Distributions; normal distribution and its properties (statements only); applications of uniform and exponential distributions; introduction to Correlation and Linear Regression.

Outcomes:

- measure of correlation between variables and obtain lines of regression

Learning Material

Review on standard probability distribution functions

There are two types of probability distributions namely (1) Discrete probability distributions (Binomial and Poisson distributions) and (2) Continuous probability distributions (Normal distribution).

Binomial distribution: Binomial distribution was discovered by James Bernoulli in the year 1700 and it is a discrete probability distribution.

Where a trial or an experiment results in only two ways say 'success' or 'failure'.

Some of the situations are: (1) Tossing a coin – head or tail (2) Birth of a baby – girl or boy (3) Auditing a bill – contains an error or not.

Definition: A random variable X is said to be Binomial distribution if it assumes only non-negative values and its probability mass function is given by

$$P(X = x) = p(x) = n_{C_x} p^x q^{n-x},\ x = 0,1,2,\ldots\ldots,n$$

$$= 0,\ otherwise$$

$where \quad q = 1 - p, \quad p + q = 1$. Here n, p are called parameters.

Example: (1) The number of defective bolts in a box containing 'n' bolts.

(2) The number of post-graduates in a group of 'n' men.

Conditions:

(1) The trials are repeated under identical conditions for a fixed number of times, say 'n' times.

(2) There are only two possible outcomes, for example success or failure for each trial.

(3) The probability of success in each trial remains constant and does not change from trail to trail.

(4) The trails are independent i.e. the probability of an event in any trail is not affected by the results of any other trail.

Constants (mean & variance) of Binomial distribution:

$$mean = E(X) = \sum xp(x)$$

$$= \sum_{x=0}^{n} x\, n_{C_x} p^x q^{n-x} = \sum_{x=1}^{n} x \frac{n}{x}\, n-1_{C_{x-i}}\, p^x q^{n-x} = np \sum_{x=1}^{n} n-1_{C_{x-1}}\, p^{x-1} q^{n-x} = np(q+p)^{n-1} = np$$

$$\text{var}\,iance = E(X^2) - [E(X)]^2$$

$$= \sum [x(x-1)+x]p(x) = \sum x(x-1)p(x) + \sum xp(x)$$

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

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

Problem: Ten coins are thrown simultaneously. Find the probability of getting at least 7 heads.

Solution: p = probability of getting a head = ½

q = probability of getting a head = ½

The p.d.f. of binomial distribution is $P(X = x) = n_{C_x} p^x q^{n-x},\ x = 0,1,2,.....,10$

Given n = 10, $P(X = x) = 10_{C_x}\, p^x q^{10-x}$

Probability of getting at least 7 heads is given by

$P(X \geq 7) = P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) = 0.1719.$

Poisson distribution: Poisson distribution due to French mathematician Denis Poisson in 1837 is a discrete probability distribution.

It is a rare distribution of rare events i.e. the events whose probability of occurrence is very small but the number of trails which would lead to the occurrence of the event, are very large.

As $n \to \infty,\ p \to 0$ B.D. tends to P.D.

Definition: A random variable X is said to follow a Poisson distribution if it assumes only non-negative values and its probability mass function is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!};\ x = 0,1,2,....$$

$$= 0,\ otherwise.$$

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

Example: (1) The number defective bulbs manufactured by a company.

(2) The number of telephone calls per minute at a switch board.

Conditions: (1) The variable or number of occurrences is a discrete variable.

(2) The occurrences are rare.

(3) The number of trails 'n' is large.

(4) The probability of success (p) is very small.

(5) np = λ is finite.

Constants (mean and variance) of Poisson distribution:

$$mean = E(X) = \sum_{x=0}^{\infty} x\, p(x) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda}.e^{\lambda} = \lambda.$$

$$\text{var}\,iance = E(X^2) - [E(X)]^2$$

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

Example: Fit a Poisson distribution for the following data and calculate the expected frequencies.

X	0	1	2	3	4

f(x)	109	65	22	3	1

Solution: By the given data, total frequency = $\sum f_i$ =200

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i} = \frac{(0)(109)+(1)(65)+(2)(22)+(3)(3)+(4)(1)}{200} = 0.61 = \lambda$$

Therefore, the theoretical frequencies = N p(x); x = 0,1,2,3,4.

i.e. $200.\frac{e^{-0.61}(0.61)^x}{x!}$ where x = 0,1,2,3,4.

When x = 0, 200 p(0) = 108.67
x = 1, 200 p(1) = 66.29
x= 2, 200 p(2) = 20.22
x= 3, 200 p(3) = 4.11
x= 4, 200 p(4) = 0.63

since frequencies are always integers, therefore by converting them to nearest integers, we get

Observed frequency	109	65	22	3	1
Expected frequency	109	66	20	4	1

Example: A car hire firm has two cars which it hires out day by day, The number of demands for a car on each day is distributed as a Poisson distribution with mean 1.5. Calculate the proportion of days (i) on which there is no demand (ii) on which demand is refused.

Solution: Given mean, λ = 1.5

We have $p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

(i) P(no demand) = p(0) = 0.2231
Number of days in a year there is no demand of car = 365 (0.2231) = 81 days.

(ii) P(demand refused) = p(x>2) = 1- [p(0)+p(1)+p(2)] = 0.1913
Number of days in a year when some demand is refused = 365 (0.1913) = 70 days.

Normal Distribution: It was first discovered by English Mathematician De-Moivre in 17333 and further refined by French Mathematician Laplace in 1744 and independently by Karl Friedrich Gauss. Normal distribution is also known as 'Gaussian distribution'.

Definition: A random variable X is said to have a Normal distribution, if its probability density function is given by

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}};\ -\infty < x < \infty,\ -\infty < \mu < \infty,\ \sigma > 0$$

Where μ is mean and σ^2 is variance are called parameters.

Notation: X~N(μ, σ^2).

Problem: In a normal distribution, 7% of the items are under 35 and 89% under 63. Determine the mean and variance of the distribution.

Solution: Given P(X<35) = 0.07 and P(X<63) = 0.89

Therefore, P(X>63) = 1- P(X<63) = 1 – 0.89 = 0.11

When X = 35, Z = (X-μ)/σ = (35-μ)/σ = -z_1 (say) ……(1)

When $X = 63$, $Z = (X-\mu)/\sigma = (63-\mu)/\sigma = -z_2$ (say) …….(2)
$P(0<Z<z_2) = 0.39 \Rightarrow z_2 = 1.23$ (from tables)
and $P(0<Z<z_1) = 0.43 \Rightarrow z_1 = 1.48$
from(1) we have $(35-\mu)/\sigma = -1.48$ ……(3)
from (2) we have $(63-\mu)/\sigma = 1.23$ ……(4)
(4) – (3) gives $\sigma = 10.332$
From equation (3), $\mu = 50.3$
Therefore, mean = 50.3 and variance = 106.75

Characteristics:
(1). The graph of the Normal distribution $y = f(x)$ in the xy-plane is known as the normal curve.
(2). The curve is a bell shaped curve and symmetrical with respect to mean i.e., about the line $x = \mu$ and the two tails on the right and the left sides of the mean (μ) extends to infinity. The top of the bell is directly above the mean μ.
(3). Area under the normal curve represents the total population.
(4). Mean, median and mode of the distribution coincide at $x = \mu$ as the distribution is symmetrical. So normal curve is unimodal (has only one maximum point).
(5). x-axix is an asymptote to the curve.
(6). Linear combination of independent normal variates is also a normal variate.
(7). The points of inflexion of the curve are at $x = \mu \pm \sigma$ and the curve changes from concave to convex at $x = \mu + \sigma$ to $x = \mu - \sigma$.
(8). The probability that the normal variate X with mean μ and standard deviation σ lies between x_1 and x_2 is given by

$$P(x_1 \le X \le x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad \text{.........(1)}$$

Since (1) depends on the two parameters μ and σ, we get different normal curves for different values of μ and σ and it is an impracticable task to plot all such normal curves. Instead, by putting $z=(x-\mu)/\sigma$, the R.H.S. of equation (1) becomes independent of the two parameters μ and σ. Here z is known as the standard variable.
(9). Area under the normal curve is distributed as follows:
$P(\mu - \sigma<X< \mu + \sigma) = 0.6826$; $P(\mu - 2\sigma<X< \mu + 2\sigma) = 0.9543$; $P(\mu - 3\sigma<X< \mu + 3\sigma) = 0.9973$.

Uniform distribution:
The uniform or rectangular distribution has a random variable X restricted to a finite interval [a,b] and has a constant over the interval.

The function f(x) can be defined as $f(x) = \begin{cases} \frac{1}{b-a}, a \le x \le b \\ 0, otherwise \end{cases}$

- The mean of the uniform distribution is $(b+a)/2$.
- The variance of the uniform distribution is $(b-a)^2/12$.

Problem: The current measured in a piece of copper wire is known to follow a uniform distribution over the interval [0, 25]. Write down the formula for the probability density function f(x) of a random variable X representing the current. Calculate the mean and variance of the distribution.
Solution:
The probability density function f(x) over the interval [0, 25] given by

$$f(x)=\begin{cases}\frac{1}{25-0}, 0 \le x \le 25\\ 0, otherwise\end{cases}$$ then Mean = 12.5 and Variance = 52.08

Exponential Distribution:

The exponential distribution of a continuous random variable X is defined as

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, \lambda > 0\\ 0, otherwise\end{cases}$$

The mean and standard deviation of an exponential distribution is 1/λ.

Problem: If x is a exponential varient with mean 5. Then find the following probabilities

i) P(0 < X < 1) ii) P(-∞ < X< 10)

Solution: Given mean is 10 then λ= 1/5.

P(X = x) = $f(x)=\begin{cases}\frac{1}{5}e^{-\frac{1}{5}x}, \lambda > 0\\ 0, otherwise\end{cases}$

i) Consider P(0 < X < 1) = $\int_0^1 f(x)dx = \frac{1}{5}\int_0^1 e^{-\frac{1}{5}x}dx = 1-\frac{1}{e^5}$.

ii) Consider P(-∞ < X< 10) = $\int_{-\infty}^{10} f(x)dx = \frac{1}{5}\int_{-\infty}^{10} e^{-\frac{1}{5}x}dx = 1-\frac{1}{e^2}$.

Correlation: It is a statistical analysis which measures and analysis the degree or extent to which two variables fluctuates with reference to each other. It expresses the relationship or independence of two sets of variables upon each other.

If change in one variable affects the change in other variable then the two vriable are said to be correlated.

Types of Correlation:

- Positive and negative
- Single and multiple
- Partial and total
- Linear and non linear

Partial and Total Correlation:

Two variables excluding some other variables is called partial correlation. Example, we study price and demand, eliminating the supply side. In total correlation, all the facts are taken into account.

Linear and non-linear correlation:

If the ratio of change between two variables is uniform, then there will be linear correlation between them.

In a curvilinear or non-linear correlation, the amount of change in one variable does not bear a constant ration of the amount of change in the other variables.

Scatter diagram or scatter gram:

The scatter diagram is pictorial representation by plotting two variables to find out whether there is any relationship between them.

Karl Pearson's correlation coefficient:

Karl Pearson is a British Biometrician and Statistician suggested a mathematical method for measuring the magnitude of linear relationship between two variables. This is known as Pearson's Coefficient of correlation or Product-Moment correlation coefficient. It is denoted by $r_{x,y}$

$$r=\frac{Cov(X,Y)}{\sigma_x\sigma_y} \quad \text{OR} \quad r=\frac{\sum XY}{\sqrt{\sum X^2\sum Y^2}} \quad \text{OR} \quad r=\frac{\sum xy}{N\sigma_x\sigma_y} \quad \text{OR} \quad r=\frac{(\sum XY*n)-(\sum X*\sum Y)}{\sqrt{(\sum X^2*n-(\sum X)^2)*(\sum Y^2*n-(\sum Y)^2)}}$$

Where n is number of paired observations

Limits of correlation coefficient ($-1 \le r_{x,y} \le +1$)

PROBLEM: Calculate coefficient of correlation from the following data.

X	12	9	8	10	11	13	7
Y	14	8	6	9	11	12	3

Solution:

We have $r=\frac{(\sum XY*n)-(\sum X*\sum Y)}{\sqrt{(\sum X^2*n-(\sum X)^2)*(\sum Y^2*n-(\sum Y)^2)}}$

X	Y	X^2	Y^2	XY
12	14	144	196	168
9	8	81	64	72
8	6	64	36	48
10	9	100	81	90
11	11	121	121	121
13	12	169	144	156
7	3	49	9	21
70	63	728	651	676

Here n=7

$\therefore r=\frac{(676*7)-(70*63)}{\sqrt{(728*7-70^2)*(651*7-63^2)}} \quad =0.95$

Note: When deviations are taken from an assumed mean the coefficient of correlation is

$$r=\frac{\sum XY-\frac{\sum X\sum Y}{n}}{\sqrt{(\sum X^2-\frac{(\sum X)^2}{n})-(\sum Y^2-\frac{(\sum Y)^2}{n})}}$$

Rank correlation coefficient:

The method of finding the coefficient of correlation by ranks. This method is based on ranks and is useful in dealing with qualitative characteristics such as morality, character, intelligence and beauty. Rank correlation is applicable only to the individual observations. The formula for Spearman's rank correlation coefficient is given by

$$\rho = 1-\frac{6\sum d^2}{n(n^2-1)} \text{ (For untied ranks)}$$

Where ρ is rank coefficient of Correlation

d^2 is Sum of the squares of the difference of two ranks

n is Number of paired observations

Properties of rank correlation coefficient:

- **T**he value of ρ lies between 1 and -1
- If ρ=1, there is complete agreement in the order if the ranks and the direction of the rank is same.
- If ρ=-1, then there is complete disagreement in the order of the ranks and they are in opposite directions.

PROBLEM: A random sample of 5 college students is selected and their grades in Mathematics and Statistics are found to be

Mathematics	85	60	73	40	90
Statistics	93	75	65	50	80

Calculate Spearman's rank correlation coefficient.

Solution:

X	Y	Ranks in x	Ranks in y	d_i =x-y	D^2
85	93	2	1	1	1
60	75	4	3	1	1
73	65	3	4	-1	1
40	50	5	5	0	0
90	80	1	2	-1	1
					4

Here N=5 $\sum D^2$=4

Spearman's rank correlation coefficient is

$$\rho = 1-\frac{6\sum d^2}{n(n^2-1)} \quad =1-\frac{6*4}{5(5^2-1)} \quad =0.8$$

Equal or Repeated ranks:

If there is more than one item with the same value in the series then the Spearman's formula for calculating the rank correlation coefficient is

$$\rho = 1\text{-}6\left\{\frac{\sum d^2 + coreection\ factor\ of\ X\ and\ Y}{n(n2-1)}\right\}$$

Where correction factor(C.F)= $m(m^2-1)/12$

Where m= the number of times the item is repeated

PROBLEM: Obtain the rank correlation coefficient for the following data

X	68	64	75	50	64	80	75	40	55	64
Y	62	58	68	45	81	60	68	48	50	70

X	Y	Rank of X(x)	Rank of Y (y)	d=x-y	d^2
68	62	4	5	-1	1
64	58	6	7	-1	1
75	68	2.5	3.5	-1	1
50	45	9	10	-1	1
64	81	6	1	5	25
80	60	1	6	-5	25
75	68	2.5	3.5	-1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16
				0	72

In X-series, 75 occurs 2 times, so rank = $\frac{2+3}{2}$ =2.5

64 occur 3 times, so rank= $\frac{5+6+7}{3} = 6$

To $\sum d^2$ we add $\frac{m(m^2-1)}{12}$ for each value repeated, so for 75 m=2, for 64, m=3.

So far X series,C.F is $\frac{2(4-1)}{12} + \frac{3(9-1)}{12} = \frac{5}{2}$

In Y series, 68 occurs twice, so rank = $\frac{3+4}{2}$ = 3.5

68 occurs twice so m=2

So far Y series, C.F is $\frac{2(4-1)}{12} = \frac{1}{2}$ $\therefore \rho = \frac{1-6(\sum d^2+\frac{5}{2}+\frac{1}{2})}{N(N^2-1)}$=0.545

Regression

In regression analysis the nature of actual relationship if it exists, between two (or more variables) is studied by determining the mathematical equation between the variables. It is mainly used to predict or estimate one (the dependent) variable in terms of the other (independent) variable(s).

Definition: Regression is a mathematical measure of the average relationship between two or more variables in terms of original units of the data.

Simple regression: It establishes the relationship between two variables (one dependent and one independent variable)

Linear regression: if the relationship between the two variables is linear and is represented by straight line then it is regression line or the line of average relationship or prediction of equation.

Regression lines are of two types (i) regression line y on x (ii) regression line x on y

The statistical method which helps us to estimate the unknown value of one variable from the known value of the related variable is called regression.

Uses:

- It is used to estimate the relation between two economic variables like income and expenditure.
- It is highly valuable tool in economic and business.
- It is useful in statistical estimation of demand curves, supply curves, production function, cost function and consumption function etc.

Properties of Regression coefficients:

1. Regression lines pass through the points (x, y)
2. Correlation coefficient is the geometric mean between the regression coefficients
3. If one of the regression coefficients is greater than unity, the other must be less than unity
4. Arithmetic mean of the regression coefficient is greater than the correlation coefficient
5. Regression coefficients are independent of the change of origin but not scale

Deviation taken from arithmetic mean X on Y:

This method is simpler to find the values of a and b. We can find out the deviations of X and Y series from their respective means.

Regression equation X on Y is

$$X-\overline{X} = r\frac{\sigma_x}{\sigma_y}(Y-\overline{Y})$$

Regression equation Y on X is

$$(Y-\overline{Y}) = r\frac{\sigma_y}{\sigma_x}(X-\overline{X})$$

Where $\overline{X}$ and $\overline{Y}$ be the means of X and Y series

The regression coefficient of X on Y = $r\frac{\sigma_x}{\sigma_y} = \frac{\sum XY}{\sum Y^2} = b_{xy}$

The regression coefficient of Y on X = $r\frac{\sigma_y}{\sigma_x} = \frac{\sum XY}{\sum X^2} = b_{yx}$

PROBLEM:

Find the most likely production corresponding to a rainfall 40 from the following data.

	Rain fall(X)	Production(Y)
Average	30	500kgs
Standard deviation	5	100kgs
Coefficient of correlation	0.8	

We have to calculate the value of Y when X =40

So we have to find the regression equation of Y on X.

Mean of X series, $\overline{X}$ =30; Mean of Y series, $\overline{Y}$ = 500
σ of X series, σ_x= 5 , σ of Y series , σ_y =100
Regression line Y on X

$$(Y-\overline{Y}) = r\frac{\sigma_y}{\sigma_x}(X-\overline{X}) = (Y-500) = 0.8\left(\frac{100}{5}\right)(X-30)$$

When X=40, Y-500=160
Y=660
Hence the expected value of Y is 660kgs.

Deviations taken from the assumed mean:

If the actual mean is fraction this method is used.
In this method we take deviations from the assumed mean instead of A.M

$$X-\overline{X} = r\frac{\sigma_x}{\sigma_y}(Y-\overline{Y})$$

We can find out the value of $r\frac{\sigma_x}{\sigma_y}$ by applying the following formula

$$r\frac{\sigma_x}{\sigma_y} = \frac{\sum dx\,dy - \frac{\sum dx * \sum dy}{n}}{\sum dy^2 - \frac{(\sum dy)^2}{n}}$$, dx =X-A; dy =Y-A

Regression equation Y on X is

$$(Y-\overline{Y}) = r\frac{\sigma_y}{\sigma_x}(X-\overline{X})$$

We can find out the value of $r\frac{\sigma_y}{\sigma_x}$ by applying the following formula

$$r\frac{\sigma_y}{\sigma_x} = \frac{\sum dx\,dy - \frac{\sum dx * \sum dy}{n}}{\sum dx^2 - \frac{(\sum dx)^2}{n}}$$

PROBLEM: Price indices of cotton and wool are given below for the 12 months of a year. Obtain the equations of lines of regression between the indices.

X	78	77	85	88	87	82	81	77	76	83	97	93
Y	84	82	82	85	89	90	88	92	83	89	98	99

Calculation of regression equation

X	dx=(X-84)	dx^2	Y	dy=(Y-88)	dy^2	Dxdy
78	-6	36	84	-4	16	24
77	-7	49	82	-6	36	42
85	1	1	82	-6	36	-6
88	4	16	85	-3	9	-12
87	3	9	89	1	1	3
82	-2	4	90	2	4	-4
81	-3	9	88	0	0	0
77	-7	49	92	4	16	-28
76	-8	64	83	-5	25	40
83	-1	1	89	1	1	-1
97	13	169	98	10	100	130
93	9	81	99	11	121	99
1004	-4	488	1061	5	365	287

Regression line X on Y:

$$X-\overline{X} = b_{xy}(Y-\overline{Y})$$

$$b_{xy} = \frac{\sum dx\,dy - \frac{\sum dx * \sum dy}{n}}{\sum dy^2 - \frac{(\sum dy)^2}{n}} = \frac{287-(\frac{-4*5}{12})}{365-\frac{5^2}{12}} = 0.795$$

X-83.7 = 0.795(Y-88.42)

X=0.795Y+13.38

Regression line Y on X:

$$(Y-\overline{Y}) = b_{yx}(X-\overline{X})$$

$$b_{yx} = \frac{\sum dx\,dy - \frac{\sum dx * \sum dy}{n}}{\sum dx^2 - \frac{(\sum dx)^2}{n}} = \frac{287-(\frac{-4*5}{12})}{488-\frac{-4^2}{12}} = 0.59$$

Y-88.42 = 0.59(X-83.67)

Y=0.59X+39.05

PROBLEM:

Determine the equation of a straight line which best fits the data.

X	10	12	13	16	17	20	25
Y	10	22	24	27	29	33	37

Let the required straight line is Y=a+bX

The two normal equations are $\sum Y$ = b$\sum X$ + na

$\sum XY$ = b$\sum X^2$+a$\sum X$

X	X^2	Y	XY
10	100	10	100
12	144	22	264
13	169	24	312
16	256	27	432
17	289	29	493
20	400	33	660
25	625	37	925
113	1938	182	3186

Substituting the values:

113b+7a =182 ------------------- (1)

1983b+113a=3186---------------- (2)

Then a=0.82, b=1.56

The equation of straight line is Y=0.82 +1.56 X

NUMERICAL AND STATISTICAL METHODS

UNIT VI

SAMPLING AND STATISTICAL INFERENCE

Objectives:

Knowing sampling theory and principles of hypothesis testing.

Syllabus:
Basic terminology in sampling, sampling techniques (with and without replacements), sampling distribution and its applications.
Introduction to statistical inference –Test for means and proportions (one sample and two samples when the sample size is large) ; Exact sample tests-Chi-Square test (Goodness of fit) and F-test (Test for population variances) ,Introduction to t-test.

Learning Outcomes:

Students should be able to
Construct sampling distribution and calculate its mean and standard deviation.
Recognize and apply the appropriate tests to give valid inference.

SAMPLING AND STATISTICAL INFERENCE

Basic Terms:
Population: In statistics population does not only refers to people but it may defined as any collection of individuals or objects or units which can be specified numerically.
Population may be mainly classified into two types.
(i) Finite population (ii) Infinite population
(i) Finite population: The population contains finite number of individuals is called 'finite population'. For example, total number of students in a class.
(ii) Infinite population: The population which contains infinite number of individuals is known as 'infinite population'. For example, the number of stars in the sky.

Parameter: The statistical constants of a population are known as parameter. For example, mean (μ) and variance (σ^2).
Sample: A portion of the population which is examined with a view to determining the population characteristics is called a sample. Or A sample is a subset of the population and the number of objects in the sample is called the size of the sample size of the sample is denoted by 'n'.
Statistic: Any function of sample observations is called sample statistic or statistic.

Standard error: The standard deviation of the sampling distribution of a statistic is known as its 'standard error'.

Classification of samples: Samples are classified in 2 ways.

(i) **Large sample:** The size of the sample (n) ≥ 30, the sample is said to be large sample.

(ii) **Small sample:** If the size of the sample (n) < 30, the sample is said to be small sample or exact sample.

Types of sampling: There are mainly 5 types of sampling as follows.

(i) **Purposive sampling:** It is one in which the sample units are selected with definite purpose in view. For example, if you want to give the picture that the standard of living has increased in the town of 'Gudivada', we may take individuals in the sample from Satyanarayana puram, Rajendra nagar etc and ignore the localities where low income group and middle class families live.

(ii) **Random sampling:** In this case sample units are selected in one in which each unit of population has an equal chance of being included in it.
Suppose we take a sample of size 'n' from finite population of size 'N'. Then there are N_{C_n} possible samples. A sampling technique in which each of the N_{C_n} samples has an equal chance of being selected is known as 'Random sampling' and the sample obtained by this is termed as 'random sample'.

(iii) **Stratified Random Sampling:** It is defined as the entire heterogeneous population is sub divided into homogeneous groups. Such groups are called 'strata'. The size of each strata may differ but they are homogeneous within themselves. A sample is drawn randomly from these strata's is known as 'stratified random sampling'.

(iv) **Systematic sampling:** In this sampling we select a random number to draw a sample and the remaining samples are automatically selected by a predetermined patterns such a sampling is called 'systematic sampling'.

(v) **Simple sampling:** Simple sampling is random sampling in which each unit of the population has an equal chance. For example, if the population consists of N units then we select a sample n units then each unit having equal probability 1/N.

Problems:

(1) Find the value of the finite population correction factor for n = 10 and N = 1000.
Given N = the size of the finite population = 1000
n = size of the sample = 10
Therefore, correction factor = (N-n)/(N-1) = 0.991

(2) A population consists of five numbers 2, 3, 6, 8 and 11. Consider all possible samples of size two which can be drawn with replacement from this population. Find
(a) The mean of the population (b) standard deviation of the population (c) mean of the sampling distribution of means (d) standard deviation of the sampling distribution of means.

Solution: (a) Mean of the population
$\mu = (2+3+6+8+11)/5 = 6$

(b) Variance (σ^2) is σ^2 = $\sum(x_i - \bar{x})^2 / n$

= $(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2$ / 5 = 10.8

Therefore, standard deviation (s.d.) $\sigma = \sqrt{10.8} = 3.29$

(c) Sampling with replacement:

Total number of samples with replacement is $N^n = 5^2 = 25$ samples of size 2. i.e.

{(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)}

Therefore, the distributin of means of the samples known as sampling distribution of means.

Therefore, the samples means are {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} and the mean of sampling distribution of means is the mean of these 25 means.

$\mu_x = (2+2.5+4+ \ldots +9.5 +11)/25 = 6.$

(d) Standard deviation:

$\sigma^2 = (2-6)^2+(2.5-6)^2+ \ldots +(9.5-6)^2+(11-6)^2 / 25 = 5.40$

Therefore , $\sigma = \sqrt{5.40} = 2.32$

(3) A population consists of 5, 10, 14, 18, 13, 24. Consider all possible samples of size 2 which can be drawn without replacement from the population. Find

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

Solution: (a) Mean of the population

$$\mu = \frac{\sum x}{n} = (5+10+14+18+13+24)/6 = 14$$

(b) Variance (σ^2) is σ^2 = $\sum(x_i - \bar{x})^2 / n$

= $(5-14)^2+(10-14)^2+\ldots+(11-6)^2$ / 6 = 35.67

Therefore, standard deviation (s.d.) $\sigma = \sqrt{10.8} = 3.29$

(c) All possible samples of size 2 i.e. the number os samples = $16_{C2} = 15$

Sample No.	Sample Values	Total of Sample values	Sample mean

1	5, 10	15	7.5
2	5, 14	19	9.5
3	5, 18	23	11.5
4	5, 13	18	9
5	5, 24	29	14.5
6	10, 14	24	12
7	10, 18	28	14
8	10, 13	23	11.5
9	10, 24	34	17
10	14,18	32	16
11	14, 13	27	13.5
12	14, 24	38	19
13	18, 13	31	15.5
14	18, 24	42	21
15	13, 24	37	18.5
			Total 210.0

(d) Variance of sampling distribution of means

$$\sigma_{\bar{X}}^2 = \frac{(7.5-14)^2 + (9.5-14)^2 + ... + (21-14)^2 + (18.5-14)^2}{15} = 14.266$$

Therefore, standard deviation, $\sigma_{\bar{X}} = \sqrt{14.266} = 3.78$

Statistical Hypothesis: Hypothesis is a statement or assumption about the population which may or may not be true

Testing of hypothesis: It is used to testing the hypothesis about the parent population from which the samples are drawn.

Test of Significance: A very important aspect of the sampling theory is the study of the test of significance, which enables us to decide on the basis of the sample results, if

- The deviation between the observed sample statistics and the hypothesis parameter value (or)
- The deviation between two independent sample statistics is significant.

Null Hypothesis: A definite statement about the population parameter. Such hypothesis which is usually a hypothesis of no difference is called 'Null hypothesis' and is usually denoted by 'H_0'.

Alternative Hypothesis: Any hypothesis which is complementary to the null hypothesis is called 'Alternative hypothesis" and is usually denoted by 'H_1'.

Eg: If we want to test the null hypothesis that the population has a specified mean μ_0 (say) i.e., H_1:μ≠ μ_0-----------------(i)

H_1:μ<μ_0------------------(ii)

H_1:μ>μ_0-------------------(iii)

Then the alternative hypothesis in (i) is known as Two-tailed test and the alternatives in (ii)and (iii) are known as left and right tailed tests respectively.

Critical Region: The region of rejection of null hypothesis H_0 when H_0 is true is that region of the outcomes at where H_0 is rejected. If the sample points falls in that region is called the 'critical region', size of the critical region is α.

Type-I error:

P(rejecting H_0 /H_0 is true) i.e., when H_0 is true it is to be accepted but it is a rejected. Therefore there is a error

Type-II error: P(Accepting H_0/H_0 is false) i.e., when H_0 is false it is to be rejected but it is accepted. Therefore there is a error

One tailed and two tailed tests: If the alternative hypothesis is of the type (< or >) and the entire critical region lies in the normal probability curve on one side then it is said to be one tailed tests (OTT)

Again the one tailed test is two types (i) Right one tailed test (ii) Left one tailed test

If the alternative hypothesis is of the type ($\neq$) and the

critical region lies in the normal probability curve on both sides then it is said to be two tailed tests (TTT)

Level of significance (LOS): The probability of committing Type-I error is known as the level of significance which is denoted by 'α'. Usually LOS are 10% , 5% or 1%.

Degrees of freedom: Suppose there are N observations and k conditions on these then the degrees of freedom is N-k. The degrees of freedom is used in small sample tests.

Procedure for testing of hypothesis:

Step (1): Set up Null hypothesis (H_0)

Step (2): Set up Alternative hypothesis (H_1) Which enables us to apply one tailed test/ Two tailed test.

Step (3): Choose Level of significance (LOS) α

Step (4): Under the null hypothesis H_0, the test statistic Z=$\frac{t-E(t)}{S.E\ of\ (t)} \sim N(0,1)$ where 't ' is a statistic

Step (5): Conclusion: If calculated Z < (tabulated) Z_α at α % LOS then accept null hypothesis otherwise reject null hypothesis.

The rejection rule for H_0: x=μ (or) μ=μ_0 is given below.

Table: Critical value of Z when n≥30

Level of significance	1%	5%	10%
Two-Tailed test	2.58	1.96	1.645
Right-Tailed test	2.33	1.645	1.28
Left-Tailed test	-2.33	-1.645	-1.28

Test of significance for a single mean: Working Rule

Step (1): Null hypothesis:$H_{0:}$μ=μ_0

Step (2): Alternative hypothesis: H_1:μ≠ μ_0/ H_1:μ<μ_0 / H_1:μ>μ_0

Step (3): Level of significance: Choose 5% (or) 1%

Step (4): Test statistic: We have the following two cases.

Case (1): When the S.D ( ) of population is known. Then the test statistic is Z = $\frac{\overline{X}-\mu}{S.E(\overline{X})} \sim N(0,1)$

Where S.E($\overline{x}$)=$\sigma/\sqrt{n}$

Where σ = standard deviation of population

n=Sample size.

Case (2): When the S.D (σ) of population is unknown. The test statistic is

$$Z=\frac{\overline{X}-\mu}{S.E(\overline{X})}$$

Where S.E $(\overline{x})=s/\sqrt{n}$

Where s = standard deviation of sample

n=Sample size.

Step (5): Conclusion: Z_{cal} is compare with Z_{tab} value.

If $Z_{cal}<Z_{tab}$ accept H_0. Otherwise reject H_0.

Problem:

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 interval for the population?

Given n=400, $\overline{x}$ =40, μ=38, $\sigma = 10$

Step (1): Null hypothesis:$H_{0:}$μ=38

Step (2): Alternative hypothesis: H_1:μ≠ 38

Step (3): Level of significance: $\alpha = 5\%$

Step (4): Test statistic: When the S.D ( ) of population is known. Then the test statistic is $Z=\frac{\overline{X}-\mu}{S.E(\overline{X})}$ Where S.E $(\overline{x})=\sigma/\sqrt{n}$

=4

Step (5): Conclusion: Z_{cal} =4, Z_{tab} =1.96

If $Z_{cal} > Z_{tab}$ at 5% LOS. So we reject H_0.

95% confidence interval is ($\overline{x}\pm 1.96\ \sigma/\sqrt{n}$) =(39.02,40.98)

Test of equality of Two means: Let $\overline{x}_1, \overline{x}_2$ be the sample means of two independent random samples sizes n_1 and n_2 drawn from two populations having the means μ_1 and μ_2 and standard deviation σ_1 andσ_2. To test whether the two population means are equal .

Step (1): Null hypothesis: $H_0:\ \mu_1=\mu_2$

Step (2): Alternative hypothesis:

$$H_1:\mu_1 \neq \mu_2 \,/\, H_1:\mu_1 \leq_2 \,/\, H_1:\mu_1 \geq \mu_2$$

Step (3): Level of significance: Choose 5% (or) 1%

Step (4): Test statistic: Z= $\dfrac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma^2}_2}{n2}}}$

Step (5): Conclusion: Z_{cal} is compare with Z_{tab} value.

If Z_{cal}<Z_{tab} accept H_0. Otherwise reject H_0.

Problem:

The mean of two large samples of sizes 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from same population of s.d 2.5 inches?

Given n_1=1000,n_2=2000 and $\overline{x_1}$ =67.5 $\overline{x_2}$ =68 population S.D σ=2.5

Step (1): Null hypothesis:$H_{0:}\mu_1=\mu_2$

Step (2): Alternative hypothesis:

$$H_1: \mu_1 \neq \mu_2$$

Step (3): Level of significance: Choose 5% (or) 1%

Step (4): Test statistic: Z= $\dfrac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma^2}_2}{n2}}}$ =5.16

Step (5): Conclusion: Z_{cal}=5.16 Z_{tab} =1.96

If Z_{cal}> Z_{tab} then we reject our H_0.

∴ The samples have not been drawn from same population of S.D 2.5 inches.

Test of significance of single proportion: Suppose a large random sample of size n has a sample proportion p of members possessing a certain attribute. To test the hypothesis that the proportion P in the population has a specified value p_0.

Step (1): Null hypothesis:$H_{0:}$p=p_0

Step (2): Alternative hypothesis: H_1:p$\neq p_0$/ H_1:p<p_0 / H_1:p>p_0

Step (3): Level of significance: Choose 5% (or) 1%

Step (4): Test statistic: Z= $\frac{p-P}{\sqrt{\frac{PQ}{n}}}$

Where p=sample proportion=x/n

P=population proportion, Q=1-P

N=sample size

Step (5): Conclusion: Z_{cal} is compare with Z_{tab} value.

If Z_{cal}<Z_{tab} accept H_0. Otherwise reject H_0.

Problem:

In a sample of 1000 people in Karnataka 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% LOS?

Given n=1000

P=sample proportion of rice eaters=540/1000=0.54

P=population proportion of rice eaters=1/2=0.5 Q=1-P=0.5

Step (1): Null hypothesis:H_0: both rice and wheat are equally popular in the state.i.,e P=0.5

Step (2): Alternative hypothesis:

H_1:p$\neq$ 0.5

Step (3): Level of significance: 1% =2.58

Step (4): Test statistic: Z= $\frac{p-P}{\sqrt{\frac{PQ}{n}}}$ =2.532

Step (5): Conclusion: Z_{cal} = 2.532, Z_{tab} =2.58.

If $Z_{cal} < Z_{tab}$ at 1% LOS then we accept H_0

Test of equality of two proportions:

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

To test whether the two samples have been drawn from the same population

Step (1): Null hypothesis: $H_0: P_1 = P_2$

Step (2): Alternative hypothesis:

$$H_1 : P_1 \neq P_2 \,/\, H_1 : P_1 \leq P_2 \,/\, H_1 : P_1 \geq P_2$$

Step (3): Level of significance: Choose 5% (or) 1%

Step (4): Test statistic: **(a)** when the population proportion P_1 and P_2 are known.

The test statistic is $Z=\frac{p_1-p_2}{S.E(p_1-p_2)} \sim N(0,1)$

Where S.E $(p_1-p_2) = \sqrt{\frac{P_1Q_1}{n_1}} + \sqrt{\frac{P_2Q_2}{n_2}}$ $\qquad Q_1=1-P_1 \qquad Q_2=1-P_2$

(b) When the population proportion P_1 and P_2 are unknown.

In this case we have two methods to estimate P_1 and P_2.

(i)Method of substitution:

In this method sample proportion p_1 and p_2 are substituted for P_1 and P_2.

$\therefore$ S.E $(p_1 - p_2) = \sqrt{\frac{p_1q_1+p_2q_2}{n_1+n_2}}$

$\therefore$ Test statistic is $Z= \frac{p_1-p_2}{S.E(p_1-p_2)}$

(ii) Method of pooling:

In this method, the estimate value for the two population proportions is obtained by pooling the two sample proportions p_1 and p_2 into a single proportion p by the formula is given below.

Sample proportion of two samples or estimated values is given by

$$P=\frac{n_1p_1+n_2p_2}{n_1+n_2}$$

$\therefore$ Test statistic is $Z=\dfrac{p_1-p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

Step (5): Conclusion: Z_{cal} is compare with Z_{tab} value.

If $Z_{cal} < Z_{tab}$ accept H_0. Otherwise reject H_0.

Problem:

In two large populations, there are 30% and 25% respectively of fair haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations?

Given n_1=1200 n_2 =900

p_1= proportion of fair haired people in first population=30/100=0.3

p_2= proportion of fair haired people in first population=25/100=0.25

Step (1): Null hypothesis: H_0: The two sample proportions are equal $p_1=p_2$

Step (2): Alternative hypothesis:

$$H_1: p_1 \neq p_2$$

Step (3): Level of significance: 5% = 1.96

Step (4): Test statistic: Z= $\dfrac{p_1-p_2}{S.E(p_1-p_2)}$ = 2.56

Step (5): Conclusion: Z_{cal}= 2.56, Z_{tab}= 1.96

If $Z_{cal} > Z_{tab}$ at 5% LOS then we reject H_0.

Problem:

(1) Random samples of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men and 325 women in favour of the proposal. Test the hypothesis that proportion of men and women in favour of the proposal at 5% LOS?

Given n_1=400 n_2 =600

p= proportion of men =200/400=0.5

p_2= proportion of women=250/600=0.541

Step (1): Null hypothesis:H_0: There is no significance difference between the option of men and women H_0: p_1=p_2=p

Step (2): Alternative hypothesis:

$$H_1: p_1 \neq p_2$$

Step (3): Level of significance: 5% = 1.96

Step (4): Test statistic: Z= P= $\frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

$$\therefore \text{ Test statistic is } Z= \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = 1.28$$

Step (5): Conclusion: Z_{cal} = 1.28, Z_{tab} = 1.96

If $Z_{cal} < Z_{tab}$ at 5% LOS then we reject H_0.

Degrees of freedom: It is very clear that in a test of hypothesis, a sample is drawn from the population of which the parameter is under test. The size of the sample varies since it depends either on the experimenter or on the resources available. Moreover, the test statistic involves the estimated value of the parameter which depends on the number of observations. Hence the sample size plays an important role in testing of hypothesis and is taken care of by degrees of freedom.

Definition: The number of independent observations in a set is called degrees of freedom.
It is denoted by ν (read as Nu). In general, the number of degrees of freedom is equal to the total number of observations less than the number of independent constraints imposed on the observations. i.e. in a set of n observations, if k is the number of independent constraints then $\nu = n - k$.

Before going to discuss the tests of significance under small samples, we need some knowledge about exact sampling distributions: t- distribution (or Student's t- distribution), F- distribution and χ^2- distribution (or Chi-Square distribution).

χ^2**- distribution:** Chi-square distribution was first discovered by Helmert in 1876 and later independently given Karl Pearson in 1900. The χ^2-distribution was discovered mainly as a measure of goodness of fit in case of frequency,

distribution, i.e. whether the observed frequencies follow a postulated distribution or not.
If $X_1, X_2, \ldots\ldots, X_n$ are n independent normal variates with mean zero and variance unity, the sum of squares of these variegates is distributed as chi-square with n degrees of –freedom.

Note:

F – Distribution as a special case of Beta dist.

χ^2 – distribution as a special case of Gamma dist.

χ^2 distribution used as non-parameter test whereas t and F distribution are parameter test.

Properties of Ch-Square distribution:

1. The χ^2 – distribution curve lies in the first quadrant since the range of X^2 is from 0 to ∞.
2. The χ^2 – distribution curve is not symmetrical and is highly positive skewed.
3. χ^2 – distribution has only one parameter v, the degrees of freedoms.
4. χ^2 –distribution curve is an unimodel curve and its mode is at the point χ^2 = (v-1).
5. The mean and variance of X^2 –distribution are v and 2v respectively.
6. The moment generating function for chi-square distribution is $M_{x^2}(t) = (1-2t)^{-v/2}$ where v = n-1.
7. Additive property holds good for any number of independent χ^2 – variates.

Application of χ^2 – test: The chi-square test is applicable

1. To test the hypothesis of the variance of population.
2. To test the goodness of fit of the theoretical distribution to observed frequency distribution, in one way classification having k-categories.
3. To test the independence of attributes, when the frequencies are presented in a two way classification (Called the contingency table) etc.,

Conditions for validity of χ^2 – test:

1. Sample size n should be large i.e. $n \geq 50$
2. If individual frequencies o (i=1,2……,n) are small say less than 10 then combine neighbouring frequencies (pooling) so that combined frequency O_i is greater than 10.
3. The number of classes' k should be independent.
4. The constraints on the cell frequencies, if any are linear.

5. The constraints on the cell frequencies, if any, are linear.

Problem: The following figures show the distribution of digits in numbers chosen at random from a telephone directory.

Digits	0	1	2	3	4	5	6	7	8	9
Frequency	1026	1107	997	966	1075	933	1107	972	964	853

Test whether the digits may be taken to occur equally frequently in the directory.

Solution:

Null hypothesis, H_0: The digits occur equally frequently in the directory.

Alternative hypothesis, H_1: The digits do not occur equally frequently under the null hypothesis, H_0 the test statistics is,

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(n-1)}$$

Where Oi is the observed frequency, Ei is expected frequency

Ei is caliculated as ∑oi /n

then expected frequency for each observed frequency Ei=10000/10 =1000

Caliculated χ^2=58.542 , χ^2 critical value at 5% LOS with 9 d.f is 16.919.

Since Caliculated $\chi^2 > \chi^2$ critical value, reject H_0

The digits do not occur equally frequently in the directory

Problem :A sample analysis of examination results of 500 students was made. It was found that 220 students had failed, 170 had secured a third class, 90 were placed in second class and 20 got a first class. Do these figures commensurate with the general examination result which is in the ratio 4 : 3 : 2 : 1 for the various categories respectively.

Solution: Null hypothesis, H_0:The observed results commensurate with the general examination results

Alternative hypothesis, H_1: The observed results do not commensurate with the general examination results

under the null hypothesis, H_0 the test statistics is,

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(n-1)}$$

Where E_i = (no. of share pertaining to Oi/total no. of shares) N

Total no. of shares=10, total frequency N=1000

No. of students who failed, O_1=220

No. of students who secured third class, O_2=170

No. of students who secured second class, O_3=90

No. of students who secured first class, O_4=20

Then E_1= (4/10)500=200, E_2= (3/10)500=150, E_3= (2/10)500=100, E_4= (1/10)500=50

Such that $E_1+E_2+E_3+E_4=500$

Calculated χ^2=23.667, χ^2 critical value at 5% LOS with 4-1=3 d.f is 7.81

Since Calculated χ^2 > χ^2 critical value, reject H_0

i.e. The observed results do not commensurate with the general examination results

Problem: Given the following contingency table for hair colour and eye colour. Find the value of χ^2? Can we expect good association between hair colour and eye colour?

		Hair colour			
		Fair	Brown	Black	Total
Eye colour	Blue	15	5	20	40
	Gray	20	10	20	50
	Brown	25	15	20	60
	Total	60	30	60	150

Null hypothesis, H_0:The two attributes hair colour and eye colour are independent

Alternative hypothesis, H_1: hair colour and eye colour are not independent

under the null hypothesis, H_0 the test statistics is,

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(m-1)\times(n-1)}$$

Where O_{ij} is observed frequency (given)

E_{ij} is expected frequency and calculated as E_{ij} =(i^{th} row total x j^{th} column total) / grand total (N)

E_{11}=(40 x 60) /150=16 similarly E_{12}=8 , E_{13}=16, E_{21}=20, E_{22}=10 , E_{23}=20, E_{31}=24,E_{32}=12, E_{33}=24

Calculated χ^2=3.6458, χ^2 critical value at 5% LOS with (3-1) (3-1)=4 d.f is 9.488

Since Calculated $\chi^2 < \chi^2$ critical value, accept H_0

i.e., Hair colour and eye colour are independent

F-distribution :- "The ratio of two sample variances is distributed of F." F-distribution was worked out by G.W. Snedecor and as a mark of respect for Sir R.A.Fisher (Father of modern statistics). Who was defined a statistics Z which is based upon the ratio of two –sample variances initially and hence it is denoted by F. (The first letter of Fisher).

Let S_1^2 be the sample variance of an independent sample of size n_1 drawn from a normal population N (μ_1, σ_1^2). Similarly , let S_2^2 be the sample variance in an independent sample of size n2 drawn from another normal population N (μ_2, σ_2^2).Thus S_1^2 and S_2^2 are the variances of two random samples of sizes n_1 and n_2 respectively drawn from two normal populations. In order to determine whether the two samples came from two populations having equal variances of the two independent random samples defined by

F= $\dfrac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} = \dfrac{\sigma_2^2 s_1^2}{\sigma_1^2 s_2^2}$

Which is an F-distribution with v1=n1-1 and v2=n2-1 degrees of freedom.

Properties of F-distribution:

(1) F-distribution curve extends on abscissa from 0 to ∞.

(2) It is an unimodel curve and its mode lies on the point F= $\dfrac{k_2(k_1-2)}{k_1(k_2+2)}$ or $\dfrac{v_2(v_1-2)}{v_1(v_2+2)}$ which is always less than unity

(3) F-distribution curve is a positive skew curve. Generally, the F-distribution curve is highly positive skewed where v2 is small

(4) The mean and variance are defined when v2 ≥ 3 and v2≥5 respectively.

(5) There exists a very useful relation for interchange of degrees of freedom v1 and v2 i.e $F_{1-\alpha}(v_1, v_2) = \dfrac{1}{F_\alpha(v_2, v_1)}$

(6) The moment generating function of F-distribution does not exist.

F-test is used to

(1)Test the hypothesis about the equality of two population variances.
(2) test the hypothesis about the equality of two or more population means.

F-test for equality of two population variances: Suppose we want to test whether two independent samples xi (i=1,2..n1) and yj(j=1,2..n2) of sizes n1 and n2 have been drawn from two normal populations with the same variance or not then

Null hypotheses, $H_0 : \sigma_x^2 = \sigma_y^2$, $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Under the null hypothesis, H_0, the test statistics is:

$$F = \frac{s_x^2}{s_y^2} \sim F_{(v_1, v_2)} \quad \text{(OR)} \quad F = \frac{s_y^2}{s_x^2} \sim F_{(v_2, v_1)}$$

When $s_x^2 > s_y^2$ OR $s_y^2 > s_x^2$ respectively

Where $s_x^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}\left(x_i - \bar{x}\right)^2$ with $\bar{x} = \frac{\sum_{i-1}^{n_1} x_i}{n_1}$

And $s_y^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}\left(y_i - \bar{y}\right)^2$ with $\bar{y} = \frac{\sum_{i-1}^{n} y_i}{n_2}$

Besides t-test, we can also apply a F-test for testing equality of two population means.

F-distribution is a very popular and useful distribution because of its utility in testing of hypothesis about the equality of several population means, two population variances and several regression coefficients in multiple regression coefficient etc.,
As a matter of fact, F-test is the backbone of analysis of variance(ANOVA)

Note: (1) F determines whether the ratio of two sample variances s1 and s2 is too small or too large.

(2) When F is close to 1, the two sample variances s1 and s2 are likely same

(3) F-distribution also known as variance ratio distribution

(4) Dividing S_1^2 and S_2^2 by their corresponding population variances standardizes the sample variance, and hence on the average both numerator and denominator approach. Therefore, its customer, to take the large sample variance as the numerator.

(5) F-distribution depends not only on the two parameters, $V_1$and V_2 but also on the order in which they are slated.

Problem: Life expectancy in 9 regions of Brazil in 1900 and in 11 regions of Brazil in 1970 was as given in the table below:

(Source: The review of income and wealth, June 1983)

Regions / Life Expectancy	1	2	3	4	5	6	7	8	9	10	11
1900	42.7	43.7	34.0	39.2	46.1	48.7	49.4	45.9	55.3	-	-
1970	54.2	50.4	44.2	49.7	55.4	57.0	58.2	56.6	61.9	57.5	53.4

It is desired to confirm, whether the variation in life expectancy in various reigns in 1900 and in 1970 in same or not.

Solution: Let the populations in 1900 and in 1970 be considered as $N(\mu_1, \sigma_1^2)$ *and* $N(\mu_1, \sigma_1^2)$ respectively.

Null hypotheses, H_0 : The variation of life expectancy in various regions in 1900 and in 1970 is same. $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Under the null hypothesis, H_0, the test statistics is :

$$F = \frac{s_1^2}{s_2^2} \sim F_{(v_1, v_2)} \quad \text{(OR)} \quad F = \frac{s_2^2}{s_1^2} \sim F_{(v_2, v_1)}$$

When $s_1^2 > s_2^2$ OR $s_2^2 > s_1^2$ respectively

Where $s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2$ with $\bar{x} = \frac{\sum_{i-1}^{n_1} x_i}{n_1}$

And $s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2$ with $\bar{y} = \frac{\sum_{i-1}^{n} y_i}{n_2}$

Calculation: $\bar{x} = \frac{405}{9} = 45, \ \bar{y} = \frac{598.5}{11} = 5441$(approximate)

$\sum_{i=1}^{9} (x_i - 45)^2$ = 5.29+1.69+121+33.64+1.21+13.69+0.9+106.9

= 288.51+19.36=302.87

$$\Rightarrow s_1^2 = \frac{302.87}{8} = 37.85$$

$\sum_{i=1}^{11}(y_i - 54.41)^2 =$ 0.04+16+104.04+22.09+1+6.76+14.44+4.84+56.25+9.61+1

=236.07

$\Rightarrow s_2^2 = \frac{236.07}{10} = 23.607$

Since $s_1^2 > s_2^2$, the value of test statistics is:

$$F = \frac{37.85}{23.607} = 1.603$$

The table value of F at 5% los with (8, 10) degrees of freedom for two tailed test is 3.85 (From F-tables).

Since F-Calculated value is less than f-tabulated value, we accept H_0. i.e. The sample data confirms the equality of variances in 1900 and 1970 in various regions of brazil or $\sigma_1^2 = \sigma_2^2$.

Practice.

Problem: The house-hold net expenditure on health care in south and north India, in two samples of households, expressed as percentage of total income is shown the following table:

South:	15.0	8.0	3.8	6.4	27.4	19.0	35.3	13.6	
North:	18.8	23.1	10.3	8.0	18.0	10.2	15.2	190.0	20.2

Test the equality of variances of households' net expenditure on health care in south and north India.

Problem: 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 of population from which these samples are drawn do not differ significantly?

Solution:

Null hypothesis, H_0: There is no significant difference between the variances of time distribution of populations. i.e. $\sigma_1^2 = \sigma_2^2$.

Alternative hypothesis, H_1: $\sigma_1^2 \neq \sigma_2^2$ (Two-tailed test)

Level of significance : Choose $\alpha = 5\% = 0.05$

Under the null hypothesis, H_0, the test statistics is :

$$F = \frac{s_1^2}{s_2^2} \sim F_{(v_1, v_2)} \quad \text{(OR)} \quad F = \frac{s_2^2}{s_1^2} \sim F_{(v_2, v_1)}$$

When $s_1^2 > s_2^2$ OR $s_2^2 > s_1^2$ respectively

Calculation: we are given $n_1 = 6$, $n_2 = 7$

$$\bar{x} = \frac{134}{6} = 22.3, \ \bar{y} = \frac{241}{7} = 34.4$$

$$\sum_{i=1}^{6}(x_i - 22.3)^2 = 81.34, \ \sum_{i=1}^{7}(y_i - 34.4)^2 = 133.72$$

$$\therefore s_1^2 = \frac{81.34}{5} = 16.26 \ and \ s_2^2 = \frac{133.72}{6} = 22.29$$

The value of test statistics is

$$F = \frac{22.29}{16.26} = 1.3699 \cong 1.37$$

F-critical value at 5% los with (5, 6) degrees of freedom for two tailed test is 4.39 (From F-tables)

Since F-Calculated value is less than F-tabulated value t 5% los, we accept H_0. i.e. there is no significant deference between the variances of the time distribution by the workers.

Problem: The nicotine contents in milligrams in two samples of tobacco were found to be as follows:

Sample A	24	27	26	21	25	-
Sample B	27	30	28	31	22	36

Can it be said that the two samples have come from the same normal population?

Hint: When testing the significance of the difference of the means of two samples, we assumed that the two samples came from the same population or from populations with same variances. If the variances of the population are not equal, a significant difference in the means may arise. Hence, to test the two samples have come from the same population or not, we need to apply both t-test and F-test. But here we note that first apply F-test, as usual manner.

t- distribution: It is discovered by W.S.Gosset in 1908. The statistician Gosset is better known by the pen name (pseudonym) ‘student’ and hence t-distribution is called student’s t-distribution.

In practice, the standard deviation σ is not known and in such a situation the only alternative left is to use S, the sample estimate of standard deviation σ. Thus, the variate $\frac{\bar{x}-\mu}{S/\sqrt{n}}$ is approximately normal provided n is sufficiently large. If n is not sufficiently large (small) the varite $\frac{\bar{x}-\mu}{S/\sqrt{n}}$ is distributed as t and hence, $t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$ where $S^2 = \frac{1}{(n-1)}\sum_i (x_i - \bar{x})^2$.

Properties of t- distribution:

(1) The shape of t- distribution is bell-shaped and is symmetrical about mean.
(2) The curve of t- distribution is asymptotic to the horizontal axis.
(3) It is symmetrical about the line t = 0.
(4) The form of the probability curve varies with degrees of freedom.
(5) It is unimodel with mean = median = mode.
(6) The mean of t- distribution is zero and variance depends upon the parameter ν, is called the degrees of freedom.
(7) The t- distribution with ν degrees of freedom approaches standard normal distribution as $\nu \to \infty$, ν being a parameter.

The t- distribution is extensively used in hypothesis about one mean or single mean, or about equality of two means or difference of means when σ is known.

Some applications of t- distribution are:

(1). To test the significance of the difference between two sample means or to compare two samples.
(2). To test the significance of an observed sample correlation coefficient and sample correlation coefficient.
(3). To test the significance of difference between two sample means or to compare two samples.
Assumptions about t- test: t- test is based on the following five assumptions.
(1). The random sample has been drawn from a population.
(2). All the observations in the sample are independent.
(3). The sample size is not large. (One should note that at least five observations are desirable for applying a t- test.)
(4). The assumed value μ_0 of the population mean is the correct value.
(5). The sample values are correctly taken and recorded.
(6).The population standard deviation σ is unknown

In case the above assumptions do not hold good, the reliability of the test decreases.

NUMERICAL AND STATISTICAL METHODS

Assignment-Cum-Tutorial Questions

UNIT-I

Algebraic and Transcendental Equations

Section-A

Objective Questions

1. The formula to find $(n+1)^{th}$ approximation of root of f(x) = 0 by Newton-Raphson method is [ ]

   a) $x_{n+1} = x_n - \dfrac{f(x)}{f(x_{n+1})}$ b) $x_{n+1} = x_n - \dfrac{f^1(x_n)}{f(x_n)}$

   c) $x_{n+1} = x_n - \dfrac{f(x_n)}{f^1(x_n)}$ d) $x_{n+1} = x_n \dfrac{f(x_n)}{f^1(x_n)}$

2. Every algebraic equation of nth degree has exactly ___________roots.
3. If root of the equation lies between a and b then f(a).f(b) _________
4. Under the conditions that f(a) and f(b) have apposite signs then the next approximation of one of the roots f(x)=0 by Regula-falsi method is given by [ ]

   a) $x_1 = \dfrac{af(a)-bf(b)}{f(a)-f(b)}$ b) $x_1 = \dfrac{af(b)-bf(a)}{f(b)-f(a)}$

   c) $x_1 = \dfrac{af(a)+bf(b)}{f(a)+f(b)}$ d) $x_1 = \dfrac{af(b)+bf(a)}{f(b)+f(a)}$

6. The interval in which the root of the equation $\boldsymbol{x^3} - 2x - 5 = 0$ lies[ ]

   a) (3,5) b) (1,2) c) (2,3) d) (5,1)

7. A root of $\boldsymbol{x^3} - 3\boldsymbol{x^2} + \mathbf{2.5} = \mathbf{0}$ lies between 1.1 and 1.2 (true or false)
8. If x_0 *and* x_1 are 1.4 and 1.5 by Regula-Falsimethod fin x_2 for $x^2 - 1 - \sin x = 0$ [ ]

   a) 1.0009 b) 1.2097 c) 1.1940 d) 1.4091

9. If first two approximations x_0 *and* x_1 of root of $x^3 - x^2 - 2 = 0$ are 1.5 and 2 then x_2 by Regula-Falsi method is [ ]

   a) 1.652 b) 1.724 c) 1.892 d) 1.928

10. If first two approximations of root of $xe^x - 3 = 0$ are 1 and 1.5 then x_2 by Regula-Falsi method is [ ]

    a) 1.21 b) 1.425 c) 1.035 d) 1.312

11. Newton's iterative formula to find the value of $3\sqrt{N}$ is [ ]

    a) $x_{n+1} = \frac{1}{3}\left(2x_n - N/x_n^2\right)$ b) $x_{n+1} = \frac{1}{3}\left(2x_n + N/x_n^2\right)$

    c) $x_{n+1} = \frac{1}{3}\left(x_n - N/x_n^2\right)$ d) $x_{n+1} = \left(2x_n - N/x_n^2\right)$

12. If first approximation x_0 for the root at $x\log_{10} x - 1.2 = 0$ is 2.74 then x_1 by Newton - Raphson method is [ ]

    a) 2.741 b)2.73 c) 2.751 d) 2.82

13. If first approximation of roots $x^2 - x - 4 = 0$ is $x_0 = 2$ then x_1 by Newton Raphsonmethod is [ ]

a) 1.82 b) 1.796 c) 1.896 d) 0.796

14. If $f(x) = x^4 - 4x - 10, x_1 =$ first approximation of root is 1.858 then x_2 by Newton-Raphson method is [ ]

a) 1.861 b) 1.872 c) 1.855 d) 1.92

Section-B

Subjective Questions

1. Write a short notes on Bisection method.
2. Explain the procedure involved in finding the solution by Regula-Falsi method.
3. Find a positive real root of $cosx + 1 - 3x = 0$ correct to two decimal places by bisection method.
4. Using Newton-Raphson method find a root of $tanx = 1.5\,x$ which is near x=1.5.
5. Find out the root of the equation$\mathbf{x^3 - x - 4 = 0}$by False Position method.
6. Find the positive root of the following equation by the method of interval halving for$x^3 - x - 1 = 0$.
7. Find a positive rootof the equation $3x = cosx + 1$ by Newton-Raphson Method.
8. Find a real root of $xe^x - \cos x = 0$ by Newton-Raphson method.
9. Find a real root of $e^x \sin x = 1$ by Regula-Falsi method.
10. Using Bisection method, find the negative root of $\mathbf{x^3}$-4x+9 =0.
11. By using the bisection method, find an approximate root of the equation $\sin x = \frac{1}{x}$ that lies between x=1 and x=1.5 (measured in radians) carry out computations up to the 7^{th} stage.
12. Find by Newton's method, a root of the equation $3x^3 - 9x^2 + 8 = 0$lying between 1 and 2 correct to three decimal places
13. Apply Newton-Raphson method find a root of $x^3 - x^2 + x - 2 = 0$ correct upto four decimal places starting from x_0 =1.
14. Find a root of the equation $x^3 - 5x + 1 = 0$ using the bisection method in 5 – stages.

Section-C

GATE/IES/Placement Test Questions

1. Newton-Raphson method is used to compute a root of the equation $x^2 - 13 = 0$ with 3.5 as the initial value. The approximation after one iteration is_______. (GATE2010)

   a) 3.575 b) 3.675 c) 3.667 d) 3.607

2. The bisection method is applied to compute a zero of the function $f(x) = x^4 - x^3 - x^2 - 4$ in the interval [1, 9]. The method converges to a solution after _________iterations. (GATE2012)

   a) 1 b) 3 c) 5 d) 7

3. The real root of the equation $5x - 2\cos x - 1 = 0$ (up to two decimal accuracy) is _______ (GATE2012)

4. Given a>0, we wish to compute Newton-Raphson iteration formula for reciprocal of a = 7 and $x_0 = 0.2$ then firs two approximations will be _______ (GATE2005)

5. Find Newton -Raphson iterative formula for $f(x) = x^2-117=0$ _______ (GATE2009)

6. If $f(x) = xe^x-2, x_0=0.8679$ then x_1 = _______________ (GATE2005)

7. If $f(x) = x^3-x^2+4x-4$, $x_0= 2$ then then x_1 = _______________ (GATE2007)

8. If $f(x) = x+\sqrt{x}-3$ and $x_0= 2$ then then x_1 = _______________ (GATE2011)

9. The Newton -Raphson iteration $x_{n+1}= \frac{x_n}{2} + \frac{3}{2x_n}$ can be used to solve the equation (GATE2002)

   a) $x^2=3$ b) $x^3=3$ c) $x^2=2$ d) $x^3=2$

10. When the Newton-Raphson method is applied to solve the equation $f(x)=x3+2x-1=0$, the solution at the end of the first iteration with the initial guess value as $x_0=1.2$ is----------------- (GATE2014)

11. Given that one root of the equation $x^3 - 10x^2 + 31x - 30 = 0$ is 5, the other two roots are_________ (GATE2007)

12. Consider the series $x_{n+1} = \frac{x_n}{2} + \frac{9}{8x_n}, x_0 = 0.5$, obtain from the Newton raphson method. The series converges to (GATE2007)

   a) 1.5 b) 1.4 c) 1.6 d) $\sqrt{2}$

13. The Bisection method is applied to compute a zero of the function $f(x) =x^4-x^3-x^2-4$ in the interval [1,9] .The method converges to a solution after --------------------- iterations (GATE2012)

14. The Newton Raphson iteration $x_{n+1} = \frac{1}{2}\left(x_n + \frac{R}{x_n}\right)$ can be used to compute the (GATE2010)

   a) Square of R b) Reciprocal of R
   c) Square root of R d) Logarithm of R

NUMERICAL AND STATISTICAL METHODS
UNIT-II
INTERPOLATION

Section-A

Objective Questions

1. The following is used for unequal interval of x values [ ]
   a) Newton's forward interpolation formula b) Lagrange's formula
   c) Newton's backward interpolation formula d) None
2. The $(n+1)^{th}$ order difference a polynomial of n^{th}degree is [ ]
   a) polynomial of n^{th} degree b) zero
   c) polynomial on first degree d) constant
3. Relation between Backward and Shifting operator is ________________
4. When do we apply Lagrange's interpolation?
5. If $y = x^2 + 2x$ then $\Delta^3 y =$_______ [ ]
   a) 1 b)2 c) 0 d)3
6. $\Delta(cosx) =$ ______________________.
7. $\frac{\delta^2}{4} + 1 =$____________________ [ ]
   a)μ b)μ^2 c)$\mu + \Delta$ d) $\Delta - 1$
8. $\left(E^{\frac{1}{2}} + E^{-\frac{1}{2}}\right)(1+\Delta)^{\frac{1}{2}} =$ ______ [ ]
   a) $1+\Delta$ b) $2+\Delta$ c)$1-\Delta$ d) Δ
9. By Newton's forward formula f(2.5)=___ for the following data[ ]

X	0	1	2
F(x)	7	10	13

   a) 15.25 b) 16.75 c) 16.25 d)16.108
10. For the following data

X	1	2	3	4
f(x)	1	4	27	64

   If x=2.5 then p= [ ]
   a) 1.5 b) 1 c) 2.5 d) 2
11. For the following data

X	0.1	0.2	0.3	0.4
f(x)	1.005	1.02	1.045	1.081

   When p=0.6, x= [ ]
   a) 0.16 b) 0.26 c) 0.1 d) 3.0 2.

Section-B

Subjective Questions

1. Certain corresponding values of x and log x are $(300, 2.4771), (304, 2.4829),\ (305, 2.4843)\ \textit{and}\ (307, 2.4871)$. Find log 301.
2. If the interval of differencing unity prove that
   $$\Delta\left(\frac{1}{f(x)}\right) = \frac{-\Delta f(x)}{f(x).f(x+1)}$$
3. Find a cubic polynomial in x which takes on the values -3, 3, 11, 27, 57 and 107, when x=0, 1, 2, 3, 4 and 5 respectively.

4. Using Newton's forward interpolation formula, for the given table of values

X	1.1	1.3	1.5	1.7	1.9
$f(x)$	0.21	0.69	1.25	1.89	2.61

Obtain the value of f(x) when x= 1.4.

5. The population of a town in the decimal census was given below. Estimate the population for the 1895

Year x	1891	1901	1911	1921	1931
Population of y	46	66	81	93	101

6. Find the cubic polynomial which takes the values
   y(0)=1,y(1)=0,y(2)=1,y(3)=10
7. Using Lagrange's interpolation, find the polynomial through (0,0),(1,1) and (2,2).
8. Find y(42) from the following data using Newton's interpolation formula

X	20	25	30	35	40	45
Y	354	332	291	260	231	204

9. Using Newton's backward formula find the value of sin38 ?

x:	0	10	20	30	40
sin x:	0	.17365	.34202	.50000	.64279

10. Using Newton's forward formula, find the value of f(1.6) if

X	1	1.4	1.8	2.2	2.6
y	3.49	4.82	5.96	6.5	8.4

11. Find log 58.75 from the following data:

X	40	45	50	55	60	65
Log x	1.60206	1.65321	1.69897	1.74036	1.77815	1.81291

Using Newton's Backward Interpolation formula.

12. Find the Lagrange's interpolating polynomial and using it find y when x = 10, if the values of x and y are given as follows:

x	5	6	9	11
y	12	13	14	16

13. The area A of a circle of diameter d is given below:

d:	80	85	90	95	100
A:	5026	5674	6362	7088	7854

Find approximately the areas of the circles of diameters 82 and 91.

NUMERICAL AND STATISTICAL METHODS
UNIT-III
Numerical Solution of ordinary Differential equations
<u>Section-A</u>

Objective Questions

1. Among the following, which is the best for solving initial value problem?
   a) Modified Euler's method b) Picard's method
   c) Runge-Kutta method of fourth order d) Taylor series method
2. If $\frac{dy}{dx}$ = f(x,y),y(x_0) = y_0,the formula for fourth order Runge – Kutta method is _______
3. If $\frac{dy}{dx}$ = -y, y(0) = 1, h = 0.01 then by Euler's method, the value of y_1 = __________
   a) 0.099 b) 0.0981 c) 0.99 d) none
4. If $\frac{dy}{dx} = \frac{y-x}{y+x}$, y(0) = 1 and h = 0.02, using Euler's method the value of y_1=________
   a) 1.02 b) 1.04 c) 1.03 d) none
5. If $\frac{dy}{dx}$ = x^2+y^2, y(0) = 0 using Taylor's series method, the value of y(0.4)=___________
   a) 0.2133 b) 0.02133 c) 0.002133 d) None
6. The value of y at x= 0.1 using Runge – Kutta method of fourth order for the differential equation $\frac{dy}{dx}$ = x – 2y, y(0) = 1 taking h = 0.1 is __________
   a) 0.825 b) 0.0825 c) 0.813 d) None
7. The value of y at x = 0.1 using modified Euler's method up to second approximation for $\frac{dy}{dx}$ = x – y, y(0) = 1 is _________
   a) 0.909 b) 0.0909 c) 0.809 d) None
8. If $\frac{dy}{dx}$ = 1 + y^2, f(x_0,y_0) =1,h – 0.2,K_1 = 0.2,K_2 = 0.202,K_3 = 0.20204,K_4 = 0.20216,then the value of y_1 by fourth order Runge – Kutta method is ________
   a) 0.0202 b) 0.202 c) 0.102 d) None
9. Using Runge – Kutta method, the approximate value of y(0.1) if $\frac{dy}{dx}$ = x + y^2,y = 1 where x =0 and f(x_0,y_0) = 1 K_1 = 0.1, K_2 = 0.115, K_3 = 0.116, K_4 = 0.134 is
   a) 1.116 b) 1.001 c) 1.211 d) None
10. Using Runge-kutta method, to solve the differential equation $\frac{dy}{dx} = x + y$, h=0.1 and y(0) = 1. The values of k_1, k_2, k_3 and k_4 respectively are
    a) 0.11, 0.121, 0.1, 0.005 b) 0.1, 0.11, 0.1105, 0.12105
    c) 0.111, 0.11105, 0.121005, 0.121 d) None of these
11. For the above problem y(0.1) = _______
    a) 1.11034 b) 1.33011 c) 1.43001 d) None of these

Section-B

Subjective Questions

1. Using Euler's method, find an approximate value of y corresponding to x=2.5. given that $\frac{dy}{dx} = \frac{x+y}{y}$ and y=2 when x=2.
2. Solve $y^1 = x-y^2$, y(0) = 1 using Taylor's series method and evaluate y(0.1), y(0.2).
3. Solve by Euler's method, $\frac{dy}{dx} = \frac{y-x}{y+x}$ given y(0)=1 and find y(1) taking stepsize h=0.5.
4. Given $y^1 = x+ \sin y$, y(0)=1 compute y(0.2) and y(0.4) with h=0.2 using modified Euler's method
5. Using R-K method, find y(0.2) for the equation $\frac{dy}{dx} = y - x$, y(0)=1,take h=0.22.
6. Using Taylor's series method, solve the equation $\frac{dy}{dx} = x^2 + y^2$ for $x = 0.4$ given that y =1 when $x = 0$ with h=0.2.
7. Find the solution of $\frac{dy}{dx}$ = x-y, y(0)=1 at x =0.1 , 0.2 ,0.3 , 0.4 and 0.5 . Using modified Euler's method
8. Using R-K method, estimate y(0.2) and y(0.4) for the equation $\frac{dy}{dx} = \frac{y^2 - x^2}{y^2 + x^2}$,y(0)=1,h=0.2
9. Find y(0.1) & y(0.2) using Euler's modified formula given that $y^1 = x^2 - y$, y(0) = 1
10. Find y(0.1) & y(0.2) using Runge - Kutta 4th order formula given that $y^1 = x^2 - y$ & y(0) =1

Section-C

GATE/IES/Placement Test Questions

1. Consider an ordinary differential equation $\frac{dx}{dt} = 4t + 4$ if *x*=0 at *t* = 0, the increment in *x* calculated using Runge-Kutta fourth order multi-step method with a step size of Δt = 0.2 is ___________ (GATE2014)

   a) 0.22 b) 0.44 c) 0.66 d) 0.88

NUMERICAL AND STATISTICAL METHODS
Assignment-cum-Tutorial Questions
Unit-IV
Probability and Expectation of Random Variable
Section-A

Objective Questions

1. When two dice are tossed, the probability of the sum of the two faces of the two dice to be 7 is __________
2. An experiment whose outcome can be predicted with maximum certainty is called ___________ experiment
3. The number of typographical errors in each page of a book is an example of __________variable
4. Given that $P(A)=0.9,\ P(B)=0.89,\ P(A\cap B)=0.84,$ then $P(A\cup B)$ is
   (a) 0.95 (b) 0.59 (c) 0.99 (d) 0.095
5. An experiment yields three mutually exclusive events A,B,C with $P(A)=2P(B)=3P(C)$ then $P(A)$ is
   (a) 2/11 (b) 3/11 (c) 6/11 (d) 5/11
6. The probability of solving a problem by the three students A, B, C respectively are 1/3, 1/4, 1/5. Then the probability that the problem will be solved is
   (a) 1/5 (b) 2/5 (c) 3/5 (d) none
7. If p_1 and p_2 are the probabilities of passing an examination respectively, then the probability of only one failing in the examination is
   (a) $p_1+p_2-2p_1p_2$ (b) p_1+p_2 (c) $p_1+p_2-p_1p_2$ (d) none
8. If two balls are drawn from a bag containing 3 white 4 black and 5 red balls, then the probability that the balls drawn are of different colors is
   (a) 47/66 (b) 10/33 (c) 5/22 (d) 2/11
9. A and $\overline{B}$ are two independent events such that $P(\overline{A}\cap B)=\dfrac{8}{25}$ and $P(A\cap \overline{B})=\dfrac{3}{25}$, then $P(A)$ is
   (a) 2/5 (b) 4/5 (c) 1/5 (d) 3/5
10. If $p(x)=x+\dfrac{2}{k}, x=1,2,3,4,5$ is the probability distribution of a discrete random variable, then $k=$
    (a) 5/7 (b) -5/7 (c) 7/5 (d) -7/5
11. If $f(x)=\dfrac{k}{(1+x^2)}, -\infty < x < \infty$ is a valid density function, then $k=$
    (a) 1/π (b) π (c) -1/π (d) none
12. If X is a continuous random variable with probability density function
    $$f(x)=\frac{(x+1)}{8},\ for\ 2<x<4$$
    $$=0, otherwise$$
    Then E(X) =

(a) 3.308 (b) 3.803 (c) 3.083 (d) 3.380

13. If X is a random variable and V(X) = 2, then V(2X + 3) =______

(a) 2 (b) 3 (c) 6 (d) 8

14. The relation between probability density function and cumulative density function of a random variable X is

(a) $F(x)=\int_{-\infty}^{x} f(x)dx$ (b) $F(x)=\int_{x}^{\infty} f(x)dx$

(c) $F(x)=\int_{-\infty}^{0} f(x)dx$ (d) $F(x)=\int_{0}^{\infty} f(x)dx$

15. If $f(x)=2e^{-2x}, x>0$ is a probability density function, then P(X ≥ 0.5) =

(a) e^{-1} (b) e^{-2} (c) e^{-3} (d) e

Section-B

Subjective Questions

1. If we draw a card from a pack, what is the probability that the card is either ace or king?
2. A die is thrown twice. What is the probability that the sum of the spots on the die at two throws is divisible by 2 or 3?
3. A bag contains 8 white and 4 red balls. One ball is drawn from the bag and it is replaced after noting its colour. In the second draw again one ball is drawn and its color is noted. What is the probability of the event that both the balls drawn are of different colours?
4. A lot of semiconductor chips have 20 defective chips. Two chips are selected at random without replacement from the lot.
   a) What is the probability that the first one selected is defective?
   b) What is the probability that the second one selected is defective, given that the first one was defective?
   c) What is the probability that both are defective?
5. If A and B are mutually exclusive events, P(A) = 0.23, and P(B) = 0.51, find

   (i) $P(\overline{A})$ (ii) $P(A\cup B)$ (iii) $P(\overline{A}\cap B)$ (iv) $P(\overline{A}\cap \overline{B})$
6. Given P(A) =0.35, P(B) = 0.73, and $P(A\cap B)=0.14$, find

   (i) $P(A\cup B)$ (ii) $P(A\cap \overline{B})$ (iii) $P(\overline{A}\cup \overline{B})$ (iv) $P(\overline{A}\cap B)$
7. A missile can be accidently launched if two relays A and B both have failed. The probabilities of A and B failing are known to be 0.01 and 0.03 respectively. It is also known that B is more likely to fail if A has failed with probability 0.06.
   (i) What is the probability of an accidental missile launch?
   (ii) What is the probability that A will fail if B has failed?
   (iii) Are the events independent?
8. A shipment of components consists of three identical boxes. One box contains 2000 components of which 25% are defective, the second box has 5000 components of which 20% are defective and the 3rd box

contains 2000 components of which 600 are defective. A box is selected at random and a component is removed at random from the box.

(i) What is the probability that this component is defective?

(ii) What is the probability that the defective component came from the second box?

9. Three machines A, B and C produce 55%, 25%, 20% of the total number of items of a factory. The percentage of defective output of these machines is 3%, 2% and 4%. If an item is selected at random, (i) find the probability that the item is defective (ii) if the selected item is defective, find the probability that the item is produced by machine A, machine B and machine C.

10. A random variable X has the following probability function value of X

X	0	1	2	3	4	5	6
P(X)	k	3k	5k	7k	9k	11k	13k

Find (i) k (ii) P(X < 4) (iii) P (X ≥ 5) (iv) P (X ≤ x) > ½?

11. Find the mean and variance of the uniform probability distribution given by f(x)=1 / n for x=1,2,.....n

12. A continuous random variable X has a p.d.f $f(x) = 4x^3$, for 0 ≤ x ≤ 1. Find the values of a and b such that (i) $P(X \le a) = P(X > a)$ (ii) $P(X > b) = 0.1$. Also find the mean and variance of the random variable X.

13. Probability density function of a random variable X is

$$f(x) = \frac{\sin x}{2},\ 0 < x < \pi$$. Find the mean and variance of the distribution.

$$= 0,\ elsewhere$$

Also calculate the probability of X lies between 0 and π/2.

14. The daily consumption of electric power (in millions of KW-hours) is a random variable having the probability density function

$$f(x) = \frac{1}{9} x e^{-x/3}, x > 0$$.

$$= 0,\ x \ge 0$$

If the total production is 12 million KW-hours, determine the probability that there is power cut (shortage) on any given day. Also find the average daily consumption of electric power.

NUMERICAL AND STATISTICAL METHODS

Unit-V
Probability Distributions and Correlation & Regression
Section-A

Objective Questions

1. If the mean of the binomial distribution is 6 and variance is 2, then p = ______________.
2. The graph of the normal curve is symmetric about the line
   (a) $x = \mu$ (b) $x = -\mu$ (c) $x = 0$ (d) $x = \pi$
3. The mean of a Poisson distribution is 8 then its variance is
   (a) 64 (b) 4 (c) 8 (d) none
4. A coin is tossed 3 times then the probability of obtaining two heads will be
   (a) 1/8 (b) 3/8 (c) 5/8 (d) 7/8
5. Mean, median and mode are equal for
   (a) Normal distribution (b) Binomial distribution
   (c) Poisson distribution (d) Bernoulli distribution.
6. For a Poisson variate, probability of getting at least one success is
   (a) $1-e^{-\lambda}$ (b) $1-e^{\lambda}$ (c) $1+e^{-\lambda}$ (d) $1+e^{\lambda}$
7. If X is a Poisson random variable such that 2 P(X = 0) = P(X = 2), then the standard deviation of X is
   (a) 2 (b) $\sqrt{2}$ (c) ½ (d) $1/\sqrt{2}$
8. In the standard normal curve the area between z = -1 and z =1 is nearly
   (a) 90% (b) 95% (c) 68% (d) 75%
9. Among the items manufactured in a factory, 5% are defective. The probability of getting one defective blade in a pack of 5 blades is
   (a) 0.2044 (b) 0.4022 (c) 0.2404 (d) 0.0244

10. If x is uniformly distributed from 5 to 12, ($5 \le x \le 12$), the mean and standard deviation of this distribution are _________.
11. If μ is equal to 4 then variance of exponential distribution is ______
12. The range of correlation coefficient is____________________
    (a) 0 t o∞ (b) -∞ to ∞ (c) 0 to 1 (d) -1 to 1
13. If the slope of the regression line is calculated to be 2.5 and the intercept 16 then the value of *Y* when *X* is 4 is________
    (a) 16 (b) 66.5 (c) 26 (d) 2.5

Section-B

Subjective Questions

1. In a test on 1000 electric bulbs, it was found that the number of bulbs was normally distributed with an average life of 2040 hours and a standard deviation of 60 hours. How many bulbs are likely to be in usage for (a) more than 2150 hours (b) less than 1950 hours (c) more than 1920 hours but less than 2100 hours?
2. Life time of IC chips manufactured by a semiconductor manufacturer is approximately normally distributed with mean 5×10^6 hours and

standard deviation of 5×10^5 hours. A mainframe manufacturer requires that at least 95% of a batch should have a lifetime greater than 4×10^6 hours. Will the deal be made?

3. In a certain junior Olympics, javelin throw distances are well approximated by a Gaussian distribution for which μ = 30m and σ = 5m. In a qualifying round, contestants must throw farther than 26m to qualify. In the main event, the record throw is 42m.
   (i) What is the probability of being disqualified in the qualifying round?
   (ii) In the main event, what is the probability that the record will be beaten?
4. Find the probabilities that a random variable having the standard normal distribution will take a value (i) between 0.87 and 1.28 (ii) between -0.34 and 0.62 (iii) greater than 0.85 (iv) greater than -0.655 along with neat diagrammatic representation.
5. A safety engineer feels that 30% of all industrial accidents in his plant are caused by failure of employees to follow instructions. If this figure is correct, find approximately, the probability that among 84 industrial accidents in the plant, anywhere from 20 to 30 (inclusive) will be due to the failure of employees to follow instructions.
6. 10% of the bolts produced by a certain machine turn out to be defective. Find the probability that in a sample of 10 bolts selected at random exactly two will be defective using (i) Binomial distribution (ii) Poisson distribution and comment on the results.
7. The number of mistakes counted in one hundred typed pages of a typist revealed that he made 2.8 mistakes on an average per page. Find the probability that (i) there is no mistake (ii) there are two or less mistakes in a page typed by him.
8. If a uniform probability distribution for the random variable X is defined with a = −2 and b = 4. Then
   (i) What is the "height" of the distribution?
   (ii) $P(X > 3)$ (iii) $P(X < 1)$ (iv) $P(X \geq 5)$ (v) $P(-1 < X < 1)$ (vi) Mean (vii) Standard Deviation
9. The time (in hours) required to repair a machine is an exponential distributed random variable with parameter $\lambda = 1/2$. What is the probability that a repair time exceeds 2 hours.
10. Calculate Karl Pearson's correlation coefficient for the following data.

X	380	402	370	365	410	392	385
Y	560	543	564	573	550	554	540

11. Price indices of cotton and wool are given below for the 12 months of a year. Obtain the equations of lines of regression between the indices.

X	36	23	27	28	28	29	30	31	33	35
Y	29	18	20	22	27	21	29	27	29	28

Estimate the Y when X is 25.

NUMERICAL AND STATISTICAL METHODS
Unit-VI
Sampling and Statistical Inference
Section-A

Objective Questions

1. The number of possible samples of size n for a population of N units with replacement is ________________________.
2. The statistical constants of the population are called
   (a) Statistic (b) Parameter (c) Sample Statistic (d) None
3. Probability of type-I error is called as ____________________
4. Whether a test is one-sided or two sided depends on _____ hypothesis.
5. A hypothesis is true, but is rejected, this is an error of type
   (a) I (b) II (c) I and II (d) None
6. Area of critical region depends on
   (a) Size of Type-I error (b) size of Type-II error
   (c) Value of the statistic (d) No. of observations
7. To test an hypothesis about proportions of items in a class, the usual test is:
   (a) t-test (b) F-test (c) Z-test (d) none of the above
8. The t-test is applicable to samples for which n is__________
9. In a t-distribution of sample size n , the degrees of freedom are ______
10. When d.f. for χ^2 are 100 or more, Chi-square is approximated to _
   (a) t-distribution (b) F-distribution
   (c) Z-distribution (d) none of the above
11. Range of variance of ratio F is :
   (a) -1to 1 (b) -∞ to ∞ (c) 0 to ∞ (d) 0 to 1
12. The shape of t—distribution is similar to that of____________
   (a) Chi-square distribution (b) F-distribution
   (c) Normal distribution (d) none
13. Which test is used to test the equality of population variances _[ ]
   (a) Chi-square test (b) t-test (c) F-test (d) z-test
14. chi-square distribution curve varies from_____________________[ ]
   (a) - ∞ to ∞ (b) - ∞ to 0 (c) 0 to ∞ (d) none

Section-B

Subjective Questions

1. A population consists of five numbers 2, 3, 6, 8 and 11. Consider all possible samples of size two which can be drawn with replacement from this population. Find (a) The mean of the population (b) standard deviation of the population (c) mean of the sampling distribution of means (d) standard deviation of the sampling distribution of means.
2. A population consists of 5, 10, 14, 18, 13and 24. Consider all possible samples of size 2 which can be drawn without replacement from the population. Find (a) mean of the population (b) standard deviation of the population (c) mean of the sampling distribution of means (d) standard deviation of the sampling distribution of means.

3. Define (a) Critical region (b) Level of significance (c) Left one tailed (d) Right one tailed. (e) Type-I and type-II errors
4. The mean life of a sample of 1000 electric bulbs produced by a company is found to be 1570hrs with a S.D of 1200hrs. If μ is the mean life time of all the bulbs produced by the company, test the hypothesis μ=1600hrs against the alternative $\mu \neq 1600$hrs at 5% LOS.
5. In a random sample of 60 workers the average time taken by them to get to work is 33.8 minutes with a S.D of 6.1 minutes. Can we reject the null hypothesis in favour of alternative hypothesis $\mu > 32.6$ at α=1% LOS.
6. Given the following information relating to two places A & B. Test whether there is any significant difference between their mean wages.

	A	B
Mean wages(Rs)	47	49
S.D(Rs)	28	40
No. of workers(Rs)	1000	1500

7. The means of two large samples of sizes 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from the same population of S.D 2.5 inches?
8. In a sample of 500 people in Tamil Nadu 280 are tea drinkers and the rest are coffee drinkers. Can we assume that both coffee and tea are equally popular in the state at 1% LOS?
9. In a big city, 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers?
10. In two large populations there are 30% and 25% of fair haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations?
11. A machine produced 20 defective articles in a batch of 400. After overhauling it produced 10 defectives in a batch of 300. Has the machine improved?
12. According to the norms established for a mechanical aptitude test, persons who are 18 years old should have average 73.2 with a S.D 8.6. If 4 randomly selected persons of that age averaged 76.7 then test the null hypothesis μ=73.2 against alternative hypothesis μ > 73.2 at the level of significance 5%.
13. The nicotine content in milligrams in two samples of tobacco were found to be as follows:

Sample A	24	27	26	21	25	----
Sample B	27	30	28	31	22	36

Can it be said that the two samples have come from the same variances?

14. In two independent samples of sizes 8 and 10 the sum of squares of deviations of the sample values from the respective sample means were 84.4 and 102.6. Test whether the difference of variances of the population is significant are not.
15. The table below shows the number of students absent on particular days in the week.

Day	Monday	Tuesday	Wednesday	Thursday	Friday
Number	125	88	85	94	108

Find the expected frequencies if it is assumed that the number of absentees is independent of the day of the week. Test, at 5% level, whether the difference in observed and expected frequencies are significant.