SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES, CHITTOOR
(AUTONOMOUS)

Approved by AICTE, New Delhi, Affiliated to JNTUA, Ananthapuramu.

LECTURE NOTES

Subject Name : Numerical & Statistical Methods

Year / Branch : II/Civil

Regulation : R23

Prepared By : Ms. N. Sandhya

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES (SITAMS)

(AUTONOMOUS)

(Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu)

Dr. D.K. Audikesavulu Marg (Bangalore-Tirupati Bye-pass Road), Murukambattu Post, CHITTOOR – 517 127, A.P.

DEPARTMENT OF SCIENCE AND HUMANITIES

NUMERICAL & STATISTICAL METHODS-(23BSC233)

Lecturer Notes prepared by
Ms. N Sandhya
Assistant Professor

II B. TECH / I SEMESTER REGULATION: R23

Branch: Civil

INDEX

YEAR & SEM: II-I

Branch: Civil

Department of Science & Humanities
II B Tech- I Semester

23BSC233	NUMERICAL & STATISTICAL METHODS	L	T	P	C
	(CIVIL)	3	1	-	4

Pre-requisite: Calculus, Differential equations, Set theory, Permutations and Combinations,

COURSE EDUCATIONAL OBJECTIVES:

1. To develop skill to analyze appropriate method to find the root of the Algebraic and Transcendental Equations and to develop skill to apply the concept of interpolation
2. To learn the method of evaluation of numerical derivative, numerical integration
3. To learn the method of evaluation of solve ordinary differential equations numerically using numerical methods.
4. To inculcate skill to investigate different applications of statistical distributions and the Corresponding conclusions required for the analysis of sample data.
5. To develop skill to apply the concept of test of significance using t-test, f-test, chi-square test, suitable of the required conclusion.

UNIT I: SOLUTION OF ALGEBRAIC & TRANSCENDENTAL EQUATIONS AND INTERPOLATION 09

Introduction-Bisection Method-Iterative method, Regula-falsi method and Newton Raphson method Finite differences-Newton's forward and backward interpolation formulae – Lagrange's formulae.

UNIT II: Numerical Differentiation, Numerical Integration and Curve Fitting 09

Numerical differentiation: Newton's forward and backward formulae
Numerical integration: Trapezoidal rule - Simpson's 1/3 Rule -Simpson's 3/8 Rule
Curve fitting: Fitting of straight line, second-degree and Exponential curve by method of least squares.

UNIT III: Solution of Initial value problems to Ordinary differential equations 09

Numerical solution of Ordinary Differential equations: Solution by Taylor's series-Picard's Method of successive Approximations-Euler's and modified Euler's methods-Runge-Kutta methods (second and fourth order).

UNIT IV: Estimation and Testing of hypothesis-large sample tests 09

Estimation-parameters, statistics, sampling distribution, point estimation, Formulation of null hypothesis, alternative hypothesis, the critical and acceptance regions, level of significance, two types of errors and power of the test. Large Sample Tests: Test for single proportion, difference of proportions, test for single mean and difference of means.

Confidence interval for parameters in one sample and two sample problems

UNITV: Testing of hypothesis-Small sample tests **09**

Student t-distribution (test for single mean, two means and paired t-test), testing of equality of variances (F-test), χ2 - test for goodness of fit, χ2 - test for independence of attributes.

COURSE OUTCOMES:

On successful completion of the course, students will be able to		Pos
CO1	Demonstrate the knowledge in solving algebraic and transcendental equations by various mathematical methods and Design novel mathematical methods for constructing the interpolating polynomials to the given data	**PO1, PO2, PO3**
CO2	Demonstrate the knowledge in finding the numerical values to derivatives, integrals through different mathematical methods	**PO1, PO2, PO3**
CO3	Demonstrate the knowledge in finding the solution of ordinary differential equations numerically through various methods	**PO1, PO2, PO3**
CO4	To identify real life problems into Mathematical Models.	**PO1, PO2, PO3**
CO5	To apply the probability theory and testing of hypothesis in the field of engineering Applications.	**PO1, PO2, PO3**

Text books:

1. S S Sastry, Introductory Methods of Numerical Analysis, PHI Learning Private Limited.
2. B. S. Grewal, Higher Engineering Mathematics, Khanna Publishers, 2017, 44th Edition
3. Miller and Freunds, Probability and Statistics for Engineers,7/e, Pearson, 2008.India.

Reference Books:

1. Erwin Kreyszig, Advanced Engineering Mathematics, John Wiley & Sons, 2018, 10th Edition.
2 R.K.Jain and S.R.K.Iyengar, Advanced Engineering Mathematics, Alpha Science International Ltd.5th Edition (9th reprint).
3 Ronald E. Walpole, Probability and Statistics for Engineers and Scientists, PNIE
4 H. K Das, Er. Rajnish Verma, Higher Engineering Mathematics, S. Chand Publications,2014, Third Edition (Reprint 2021)

Online Learning Resources:

1. https://onlinecourses.nptel.ac.in/noc17_ma14/preview
2. https://onlinecourses.nptel.ac.in/noc24_ma05/preview
3. http://nptel.ac.in/courses/111105090

CO\PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO.1	3	3	3	-	-	-	-	-	-	-	-	-
CO.2	3	3	3	-	-	-	-	-	-	-	-	-
CO.3	3	3	3	-	-	-	-	-	-	-	-	-
CO.4	3	3	3	-	-	-	-	-	-	-	-	-
CO.5	3	3	3	-	-	-	-	-	-	-	-	-
CO*	**3**	**3**	**3**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	-

🎓 Chapter 0: Introduction to Numerical and Statistical Methods

💡 Theme

"When Mathematics Meets Real Life"

Opening Activity

Question

Can a computer solve every mathematical problem exactly?

Examples:

- **√2 = 1.414213...**
- **π = 3.141592...**
- **Weather prediction**
- **Google Maps shortest route**
- **AI predictions**
- **Cricket win probability**

Answer: Mostly No

Conclusion

When exact solutions are difficult or impossible, we use Numerical Methods and Statistical Methods.

🚀 What is Numerical Methods?

Definition

Numerical Methods are mathematical techniques used to obtain approximate solutions to engineering and scientific problems when exact analytical solutions are impossible or difficult to obtain.

📌 Simple Definition

Numerical Methods = Approximate Solutions + Computer Algorithms

🚀 Statistical Methods

A branch of mathematics used to collect, organize , analyze , interpret, and draw conclusions from data.

Why Should Engineers Study This Course?

Engineering Problem	Method Used
Finding roots of equations	Numerical Methods
Weather Forecasting	Numerical Methods
AI & Machine Learning	Statistics
Data Analysis	Statistics
Structural Design	Numerical Methods
Medical Diagnosis	Statistics
Business Forecasting	Statistics
Robotics	Numerical Methods

Real-Life Examples

Problem	Numerical Method Used
Google Maps shortest route	Optimization Algorithms
Weather Forecasting	Numerical Differential Equations
ISRO Rocket Launch	Newton-Raphson Method
Bridge Design	Numerical Integration

Real-Life Applications

Civil Engineering

- Bridge design
- Load calculations
- Water resource modelling

Mechanical Engineering

- Heat transfer analysis
- Engine design
- Fluid mechanics

Computer Science

- Artificial Intelligence
- Machine Learning
- Data Science

Electronics

- Signal processing
- Image processing

Electrical Engineering

- Power system analysis
- Optimization

Course Roadmap

Unit I

Solution of Algebraic & Transcendental Equations

- Bisection Method
- Iteration Method
- Regula-Falsi Method
- Newton-Raphson Method
- Interpolation

Unit II

Numerical Differentiation

Numerical Integration

Curve Fitting

Applications:

- Area calculation
- Speed estimation
- Data modelling

Unit III

Ordinary Differential Equations

Methods:

- Taylor's Method

- Picard's Method
- Euler's Method
- Modified Euler's Method
- Runge-Kutta Method

Applications:

- Population growth
- Electric circuits
- Motion of objects

Unit IV
Large Sample Tests

Topics:

- Sampling
- Estimation
- Hypothesis Testing
- Z-Test
- Confidence Intervals

Applications:

- Surveys
- Manufacturing quality
- Public health

Unit V
Small Sample Tests

Topics:

- t-Test
- F-Test
- Chi-Square Test

Applications:

- Medical research
- Product comparison
- Scientific experiments

Course Outcomes

By the end of this course, students will be able to:

- ☑ **Solve algebraic and transcendental equations numerically.**
- ☑ **Approximate derivatives and integrals using numerical techniques.**
- ☑ **Solve ordinary differential equations using numerical methods.**
- ☑ **Analyze real-world data using statistical techniques.**
- ☑ **Perform hypothesis testing and make data-driven decisions.**

Mind Map

NUMERICAL METHODS

|

Root Finding | **Numerical Integration** | **Differential Equations**

|

Testing of Hypothesis for large samples

|

Testing of Hypothesis for Small samples

|

Engineering Applications

Did you know? ◇

- Google Maps updates every second using numerical computations.
- ChatGPT learns from huge amounts of statistical data.
- Hospitals use hypothesis testing before approving medicines.
- IPL teams use statistics to select players.
- Banks predict loan repayment using statistical models.
- Self-driving cars continuously solve numerical equations to avoid

obstacles.

ISRO uses Numerical Methods for satellite trajectory calculations.

- **Google Maps uses optimization algorithms to suggest the fastest route.**
- **Weather forecasting uses millions of numerical calculations every second.**
- **Artificial Intelligence is built on numerical optimization techniques.**

💬 Classroom Discussion

Think & Answer

1. **Can every mathematical equation be solved exactly?**
2. **Why do engineers prefer numerical methods over analytical methods?**
3. **Why are computers essential for Numerical Methods?**

✨ Key Takeaways

- **Numerical Methods provide approximate solutions to complex engineering problems.**
- **They are essential for solving real-world problems that lack exact analytical solutions.**
- **These techniques are widely used in mechanical, civil, electrical, electronics, computer science, aerospace, biomedical, and data science applications.**

✸ Inspirational Quote

"Engineering begins where exact mathematics ends. Numerical Methods bridge the gap between theory and real-world engineering solutions."

UNIT-1

SOLUTION OF ALGEBRAIC AND TRANSCENDENTAL

UNIT-1
SOLUTION OF ALGEBRAIC AND TRANSCENDENTAL EQUATIONS

Introduction:

The determination of the roots (zeros) of an equation of the form

$$f(x) = 0$$

is an important problem in science, engineering, economics, and mathematics. Many equations cannot be solved exactly, so numerical methods are used to obtain approximate solutions.

🎯 Learning Outcomes

After this class, students will be able to:

- **Differentiate between algebraic and transcendental equations.**
- **Understand the concept of the root of an equation.**
- **Recognize the need for numerical methods.**
- **Identify real-world engineering applications.**

🤔 Warm-Up Activity (Think–Pair–Share)

Question 1

Solve:

$$2x + 5 = 9$$

Answer:

$$x = 2$$

Easy! ✓

Question 2

Now solve:

$$x = \cos x$$

Can you solve it using ordinary algebra?
No!
This is where Numerical Methods come into the picture.

📌 What is an Equation?
An equation is a mathematical statement containing one or more unknown variables.
Example:

$$2x + 3 = 7$$

📌 What is the Root of an Equation?
A root (or zero) of an equation is the value of x that satisfies the equation.
Example:

$$x^2 - 9 = 0$$

Roots are:

$$x = 3, x = -3$$

Importance of Finding Roots:

Finding roots is essential in many fields such as:

- **Engineering design**
- **Electrical circuit analysis**
- **Structural analysis**
- **Fluid mechanics**
- **Economics**
- **Computer science**
- **Physics**

Many practical problems lead to equations that cannot be solved analytically. Numerical methods provide approximate solutions with the desired accuracy.

Types of Equations

1 Algebraic Equation

An equation containing only algebraic operations such as addition, subtraction, multiplication, division, powers, and roots of the variable.

Examples

$$x^2 - 5x + 6 = 0$$
$$x^3 - 4x + 1 = 0$$
$$x^4 + 2x^2 - 7 = 0$$

OR

A polynomial equation of the form

$$f(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \cdots + a_{n-1}x + a_n = 0$$

where $a_0, a_1, \dots, a_n$ are constants and $a_0 \neq 0$, is called an algebraic equation.
Examples : $x^4 - 5x^2 + 6 = 0$, $x^2 + x + 1 = 0$

Characteristics

✓ Polynomial equations
✓ Finite number of terms
✓ Can often be solved exactly (for lower degrees)

---

2 Transcendental Equation

An equation involving one or more transcendental functions such as:

- **Trigonometric functions**
- **Exponential functions**
- **Logarithmic functions**
- **Hyperbolic functions**

is called a transcendental equation.

Examples

$$\tan x - e^x = 0$$
$$xe^x = \cos x$$

$$\sin x = xe^{2x}$$

These equations usually cannot be solved exactly, so numerical methods are required

Characteristics

✓ Cannot usually be solved exactly.
✓ Numerical methods are required.

Comparison

Algebraic Equation	Transcendental Equation
Polynomial equation	Contains exponential, logarithmic, or trigonometric functions
Exact solution may exist	Exact solution is generally not possible
Easier to solve	Requires numerical methods

Important Terms

Equation: A mathematical statement with an unknown variable.
Root (Zero): The value that satisfies the equation.
Approximate Solution: A value close to the exact solution.
Iteration: Repeating calculations until the required accuracy is achieved.

Quick Quiz

1. What is the root of an equation?
2. Which equation is transcendental?

a) $x^2 - 5 = 0$
b) $e^x - x = 0$

3. Which engineering field uses numerical methods?
a) Mechanical
b) Civil
c) Electrical
d) All of the above
Answer: d) All of the above

📑 Key Points to Remember

- A root is the value that satisfies an equation.
- Equations are classified as algebraic and transcendental.
- Transcendental equations generally require numerical methods.
- Numerical methods provide accurate approximate solutions.

✸ Activity

Find whether the following equations are Algebraic or Transcendental:

1. $x^3 - 2x + 1 = 0$
2. $e^x - x = 0$
3. $\sin x = x/2$
4. $x^4 + 5x - 3 = 0$
5. $\log x + x = 2$

✨ Motivational Quote

"Every complex engineering problem begins with an equation, and every successful engineer knows how to solve it."

Graphical Method and Bracketing of Roots

🎯 Case Study:

Designing a Water Storage Tank

🌍 Real-Life Scenario

A civil engineer is designing a water storage tank. During the design process, the engineer needs to determine the height of water (h) that satisfies the equation:

$$f(h) = h^3 - 6h - 4 = 0$$

Since this equation cannot be solved easily using ordinary algebra, the engineer first decides to identify an interval where the solution exists before applying a numerical method.

🔍 Step 1: Check the Interval

Choose the interval:

$$h = 2 \text{and} h = 3$$

Calculate the function values:

$$f(2) = 2^3 - 6(2) - 4 = 8 - 12 - 4 = -8$$
$$f(3) = 3^3 - 6(3) - 4 = 27 - 18 - 4 = 5$$

📌 Observation

Since

$$f(2) \times f(3) = (-8)(5) = -40 < 0$$

the function changes its sign between 2 and 3.
Therefore,

☑ At least one root lies in the interval (2, 3).
This interval is called the Bracketing Interval.

📈 Graphical Interpretation

If the graph of $f(h) = h^3 - 6h - 4$is plotted,

- **At h = 2, the graph is below the x-axis.**
- **At h = 3, the graph is above the x-axis.**

Hence, the graph crosses the x-axis between 2 and 3, indicating the presence of a root.

💡 Engineer's Decision

The engineer successfully identifies the interval (2, 3) containing the root. The next step is to apply a numerical method such as the Bisection Method or Regula-Falsi Method to determine the exact height of the water.

🌍 Engineering Significance

Bracketing of roots is widely used in:

- 🏗 **Structural Engineering – Load and stress analysis.**
- 🚜 **Mechanical Engineering – Equilibrium and vibration analysis.**
- ⚡ **Electrical Engineering – Circuit parameter calculations.**
- 🌊 **Civil Engineering – Water resource and hydraulic system design.**
- 🚀 **Aerospace Engineering – Flight trajectory calculations.**

💬 Classroom Discussion

1. **Why is it important to identify a bracketing interval before using the Bisection Method?**
2. **What condition guarantees the existence of a root in an interval?**
3. **Can the Graphical Method provide the exact value of a root? Why or why not?**
4. **Which numerical methods require a bracketing interval?**

🎯 Learning Outcome

By the end of this case study, students will understand that Graphical Method and Bracketing of Roots are essential first steps in solving nonlinear engineering equations. These techniques help engineers locate the interval containing the solution before applying numerical methods to obtain an accurate root.

Takeaway: *"Before finding the solution, an engineer must first know where the solution exists."*

✹ Final Inspirational Message

"Numerical Methods are the bridge between mathematical theory and real-world engineering solutions. Every modern engineer uses them to solve problems that shape the future."

Numerical Methods Used to Find Roots

The common methods are:

1. **Bisection Method**
2. **Regula-Falsi (False Position) Method**
3. **Newton-Raphson Method**
4. **Secant Method**
5. **Fixed Point Iteration Method**

Geometrical Interpretation

(a) Simple Root

A number $\alpha$is called a **simple root** if $f(\alpha) = 0$
And $f'(\alpha) \neq 0$.
The function can be written as$f(x) = (x - \alpha)g(x)$,
where

$$g(\alpha) \neq 0.$$

Geometrical Meaning:

- The graph **crosses the x-axis** at the root.
- The slope at the root is **not zero**.

Example:

$$f(x) = x - 3.$$

The graph cuts the x-axis at

$$x = 3.$$

(b) Multiple Root:

A number $\alpha$is called a **multiple root of multiplicity** $m$if

$$f(\alpha) = f'(\alpha) = \cdots = f^{(m-1)}(\alpha) = 0$$

but

$$f^{(m)}(\alpha) \neq 0.$$

The function can be written as

$$f(x) = (x - \alpha)^m g(x),$$

where

$$g(\alpha) \neq 0.$$

Geometrical Meaning:

- If m is **even** (2, 4, 6, ...), the graph **touches the x-axis and turns back**.
- If m is **odd** (3, 5, ...), the graph **crosses the x-axis**, but more flatly than a simple root.

Example:

$$f(x) = (x-2)^2.$$

Here,

- Root = 2
- Multiplicity = 2

The graph touches the x-axis at $x = 2$and turns back.

Difference Between Simple and Multiple Root:

Simple Root	Multiple Root
$f(\alpha) = 0$	$f(\alpha) = f'(\alpha) = \cdots = f^{(m-1)}(\alpha) = 0$
$f'(\alpha) \neq 0$	First non-zero derivative is the m-th derivative
Graph crosses x-axis	Even multiplicity: touches x-axis; Odd multiplicity: crosses x-axis
Multiplicity = 1	Multiplicity > 1

Number and Methods of Finding Roots:

(a) Number of Roots:

Polynomial Equations:

A polynomial of degree $n$has exactly n**roots** (counting repeated and complex roots).

Example

$$x^3 - 6x^2 + 11x - 6 = 0$$

has 3 roots:

$$1,\ 2,\ 3.$$

Transcendental Equations:

A transcendental equation may have:

- One root

- No root
- Two or more roots
- Infinitely many roots

depending on the function.

Example

$$\sin x = 0$$

has infinitely many roots:

$$x = n\pi, n = 0, \pm 1, \pm 2, \dots$$

Methods of Finding Roots:

1. Direct Method

- Gives the **exact value** of the root.
- Solves the equation in a finite number of steps.
- Suitable for simple algebraic equations.

Example:

$$x^2 - 9 = 0$$

Roots: $x = \pm 3$.

---

2. Numerical Method:

Numerical methods are used when the equation cannot be solved exactly.
Starting with an initial approximation

$$x_0,$$

successive approximations

$$x_1, x_2, x_3, \dots$$

are obtained until they converge to the actual root

$$\alpha.$$

Mathematically,

$$x_k \to \alpha ask \to \infty.$$

Examples of numerical methods:

- Bisection Method
- Regula-Falsi Method
- Newton-Raphson Method
- Secant Method
- Fixed Point Iteration Method

- **Root:** A value α such that $f(\alpha) = 0$.
- **Geometrical Interpretation:** Root is the point where the graph meets the x-axis.
- **Simple Root:** $f(\alpha) = 0,\ f'(\alpha) \neq 0$; graph crosses the x-axis.
- **Multiple Root:** $f(\alpha) = f'(\alpha) = \cdots = f^{(m-1)}(\alpha) = 0$; even multiplicity touches the x-axis, odd multiplicity crosses it.
- **Polynomial of degree n:** Has exactly $n$roots.
- **Numerical methods:** Used to find approximate roots when exact solutions are not possible.

1. Bisection Method (Bolzano Method):

Definition

The **Bisection Method** is used to find the approximate root of an equation

$$f(x) = 0$$

by repeatedly dividing the interval into two equal parts.
This method works only if the function is **continuous** in the interval $[a, b]$and

$$f(a) \times f(b) < 0.$$

This means:

- $f(a)$is negative and $f(b)$is positive, or
- $f(a)$is positive and $f(b)$is negative.

Since the signs are opposite, there is at least one root between $a$and b.

Algorithm (Procedure)

Step 1

Choose two values $a$and $b$such that

$$f(a) \times f(b) < 0.$$

Step 2

Find the midpoint

$$x_1 = \frac{a+b}{2}.$$

This is the **first approximation**.

Step 3

Calculate

$$f(x_1).$$

- If $f(x_1) = 0,$ then x_1 is the exact root. Otherwise continue.

Step 4 If

$$f(a) \times f(x_1) < 0,$$

then the root lies between

$$a \text{ and } x_1.$$

Take the new interval as

$$[a, x_1].$$

Otherwise,
if

$$f(x_1) \times f(b) < 0,$$

the root lies between

$$x_1 \text{ and } b.$$

Take the new interval as

$$[x_1, b].$$

Step 5

Again find the midpoint of the new interval.
For example,
if the new interval is

$$[a, x_1],$$

then

$$x_2 = \frac{a + x_1}{2}.$$

Step 6
Repeat the same process until the required accuracy is obtained.

Formulae:

First Approximation

$$x_1 = \frac{a + b}{2}$$

Second Approximation

If the root lies in $[a, x_1]$,

$$x_2 = \frac{a + x_1}{2}$$

If the root lies in $[x_1, b]$,

$$x_2 = \frac{x_1 + b}{2}$$

Similarly,

$$x_3, x_4, \dots$$

are obtained by repeatedly bisecting the interval.

Flow of the Method

Choose a and b

|
▼
Check f(a)×f(b)<0 ?

|
▼
Find midpoint

$x_1=(a+b)/2$

|
▼
Calculate $f(x_1)$

|
▼
Choose the interval where sign changes

|
▼
Repeat until required accuracy

Geometrical Interpretation

```
↑y
|
|          • (b,f(b))
|         /
|        /
|       /
|      /
|-------●-------------→ x-axis
|    Root
|
|  • (a,f(a))
|+--------------------------------→ x
```

f(a)<0 f(b)>0

The graph crosses the x-axis between **a** and **b**.
The midpoint

$$x_1 = \frac{a+b}{2}$$

is calculated.
If $f(x_1)$is positive, the new interval becomes

$$[a, x_1].$$

If $f(x_1)$is negative, the new interval becomes

$$[x_1, b].$$

🕮 Case Study

Engine Safe Operating Temperature

The temperature satisfies

$$T^3 - 4T - 9 = 0$$

Checking

$$f(2) = -9$$
$$f(3) = 6$$

Since

$$(-9)(6) < 0$$

the engineer applies the Bisection Method.

After several iterations, the safe operating temperature is obtained accurately

Advantages:

- Very simple and easy to understand.
- Always converges if $f(a)$and $f(b)$have opposite signs.
- Gives accurate results with repeated iterations.
- Suitable for continuous functions.

Disadvantages:

- Converges slowly compared to other numerical methods.
- Applicable only when the initial interval has opposite signs.
- Cannot be used if $f(a) \times f(b) > 0$.

Important Conditions:

The Bisection Method can be applied only when:

1. The function is continuous.
2. The interval $[a, b]$is known.
3. $f(a) \times f(b) < 0$.

Exam Points (2 Marks)

- **Bisection Method:** A numerical method that finds the root by repeatedly dividing the interval into two equal parts.
- **Also called: Bolzano Method**.
- **First approximation:**

$$x_1 = \frac{a+b}{2}$$

- **Condition for applicability:**

$$f(a) \times f(b) < 0$$

- **Main idea:** Select the subinterval in which the sign of the function changes and continue bisecting until the desired accuracy is reached.

PROBLEMS

Problem 1: Find the Real Root of $x^3 - x - 1 = 0$by Bisection Method

Given

$$f(x) = x^3 - x - 1$$

Step 1: Find the Interval

Calculate the values of the function.

$$f(0) = 0 - 0 - 1 = -1 < 0$$
$$f(1) = 1 - 1 - 1 = -1 < 0$$
$$f(2) = 8 - 2 - 1 = 5 > 0$$

Since

$$f(1) < 0, f(2) > 0,$$

the root lies between **1 and 2**.

So,

$$a = 1, b = 2.$$

Step 2: Apply Bisection Formula

The formula is

$$x_n = \frac{a+b}{2}.$$

At each step:

- Find the midpoint.
- Calculate $f(x_n)$.

- Choose the interval where the sign changes.
- Repeat until the required accuracy is obtained.

Iteration Table

Iteration (n)	A	B	$x_n = \frac{a+b}{2}$	$f(x_n)$	Sign
1	1	2	1.500	0.875	+ve
2	1	1.500	1.250	-0.297	-ve
3	1.250	1.500	1.375	0.225	+ve
4	1.250	1.375	1.3125	-0.052	–ve
5	1.3125	1.375	1.34375	0.082	+ve
6	1.3125	1.34375	1.328125	0.014	+ve
7	1.3125	1.328125	1.3203125	-0.019	–ve
8	1.3203125	1.328125	1.32421875	0.001	+ve
9	1.3203125	1.32421875	1.322265625	-0.009	–ve
10	1.322265625	1.32421875	1.3232421875	-0.004 (approx.)	–ve
11	1.3232421875	1.32421875	1.32373 (approx.)	≈0	

Result

Therefore, the required real root is

$$\boxed{x \approx 1.322}$$

Problem 2: Find the Real Root of $xe^x - 1 = 0$ by Bisection Method

Given

$$f(x) = \mathbf{x}\boldsymbol{e}^{\boldsymbol{x}} - \mathbf{1}$$

Solution:

Step 1: Find the Interval

Calculate the values of the function.

$$f(0) = \mathbf{0}\boldsymbol{e}^{\mathbf{0}} - \mathbf{1} = -1 < 0 \quad \textit{(-ve)}$$

$$f(1) = \mathbf{1}\boldsymbol{e}^{\mathbf{1}} - \mathbf{1} = 1.7183 > 0 \quad \textit{(+ve)}$$

Since

$$f(0) < 0, f(1) > 0,$$

the root lies between **0 and 1**.
So,

$$a = 0, b = 1.$$

Table representing Bisection Method iterations:

N	a (-ve)	b (+ve)	xn = (a+b)/2	f(xn) = xn*e^xn - 1	Sign of f(xn)
1	0	1	0.5	-0.1756	-ve
2	0.5	1	0.75	0.5878	(+ve)
3	0.5	0.75	0.625	0.1677	(+ve)
4	0.5	0.625	0.5625	-0.0128	(+ve)
5	0.5625	0.625	0.5938	0.0751	+ve
6	0.5625	0.5938	0.5782	0.0307	(+ve)
7	0.5625	0.5782	0.5704	0.0089	(+ve)
8	0.5685	0.5704	0.5685	-0.0019	(-ve)
9	0.5665	0.5704	0.5685	0.0036	(+ve)
10	0.5665	0.5685	0.5675	0.0010	(+ve)
11	0.5665	0.5675	0.567	-0.0004	(-ve)
12	0.567	0.5675	0.5673	0.0003	(+ve)
13	0.567	0.5673	0.5672	0.0002	(+ve)
14	0.567	0.5672	0.5671	-0.0001	(-ve)
15	0.567	0.5672	0.5672	0.00002	(+ve)
16	0.5671	0.5672	0.5672	0.00002	(+ve)

The root is 0.5672

3. Find a real root of the equation f(x) = x^3 - 4x - 9, using Bisection Method

Let f(x) = x^3 - 4x - 9

put x=0 => f(0) = -9 (-ve)
put x=1 => f(1) = -12 (-ve)
put x=2 => f(2) = -9 (-ve)
put x=3 => f(3) = 6 (+ve)
.. f(2) & f(3) have opposite signs
.. root lies b/w 2 & 3
Let xn = (a+b)/2

N	a (-ve)	b (+ve)	xn = (a+b)/2	f(xn) = x^3-4x-9	Sign
0	2	3	2.5	-3.3150	-ve
1	2.5	3	2.75	0.7969	+ve
2	2.5	2.75	2.6250	-1.4121	-ve
3	2.625	2.75	2.6875	-0.3371	-ve
4	2.6875	2.75	2.7188	0.2209	+ve
5	2.6875	2.7188	2.7032	-0.0606	-ve
6	2.7032	2.7188	2.7110	0.0806	+ve
7	2.7032	2.7110	2.7071	0.0103	+ve
8	2.7032	2.7071	2.7052	-0.0248	-ve
9	2.7052	2.7071	2.7062	-0.0068	-ve
10	2.7062	2.7071	2.7067	0.0022	+ve
11	2.7062	2.7067	2.7065	-0.0014	-ve
12	2.7065	2.7067	2.7066	0.0013	+ve
13	2.7065	2.7066	2.7066	0.004	+ve

.. x12 = x13
.. The root is 2.7066

2. Regula Falsi Method (or) False Position Method:

The given equation is f(x)=0. Suppose we want to find the roots of equation f(x)=0 first we select the interval [a,b] such that f(a) & f(b) have the approximation signs then root of f(x)=0 lies b/w a & b

Then find first approximation x1 = (a*f(b) - b*f(a)) / (f(b) - f(a))
If f(x1) = 0, then x1 is root
If f(x1) is +ve and f(a) is -ve then root lies b/w a & x1
Now 2nd approximation is x2 = (a*f(x1) - x1*f(a)) / (f(x1) - f(a))
Suppose f(x1) & f(b) are opposite signs then the roots lies b/w x1 & b.
Now x3 = (x1*f(b) - b*f(x1)) / (f(b) - f(x1))

Advantages of Regula Falsi (False Position) Method

1. It is **simple and easy to understand**.
2. It **always converges** if the function is continuous and the initial interval brackets the root (i.e., $f(a)$and $f(b)$have opposite signs).
3. It does **not require the derivative** of the function.
4. It generally converges **faster than the Bisection Method**.
5. The root always remains **within the selected interval**, making it a reliable method.

Disadvantages of Regula Falsi (False Position) Method

1. The **rate of convergence may be slow**, especially when one endpoint remains fixed for many iterations.
2. It may require **many iterations** to obtain the desired accuracy.
3. It is applicable **only when the root is bracketed** (i.e., the function changes sign over the interval).
4. It may converge **more slowly than methods like the** Newton–Raphson Method **or** Secant Method.
5. The method is **not suitable for functions that do not change sign** over the chosen interval, even if a root exists.

Problems:

1. Find the root of the equation x^3 - 2x - 5 = 0

let f(x) = x^3 - 2x - 5
put x=0 => f(0) = -5 (-ve)
put x=1 => f(1) = 1 - 2 - 5 = -6 (-ve)
put x=2 => f(2) = 2^3 - 4 - 5 = -1 (-ve)
put x=3 => f(3) = 3^3 - 6 - 5 = 16 (+ve)
.. f(2) & f(3) have opposite signs
.. Root lies b/w 2 & 3

In Regula falsi method $x_n = \frac{af(b) - bf(a)}{f(b) - f(a)}$

N	A	B	f(a)	f(b)	$x_n = \frac{af(b) - bf(a)}{f(b) - f(a)}$	f(x_n) = x^3-2x-5
1	2	3	-1	16	2.0588	-0.390

2	2.0588	3	-0.3908	16	2.0813	-0.147
3	2.0813	3	-0.1470	16	2.0896	-0.054
4	2.0896	3	-0.0546	16	2.0927	-0.020
5	2.0927	3	-0.0202	16	2.0939	-0.007
6	2.0939	3	-0.0075	16	2.0943	-0.002
7	2.0943	3	-0.0028	16	2.0945	-0.0010
8	2.0945	3	-0.0010	16	2.0946	-0.00037
9	2.0946	3	-0.00037	16	2.0946	-0.00014

.. x8 = x9

.. The root is 2.0946

2. Solve x*logx - 1.2 = 0 using Regula falsi method

let f(x) = x*logx - 1.2

put x=1 => f(1) = -1.2 < 0 (-ve)

put x=2 => f(2) = -0.5979 < 0 (-ve)

put x=3 => f(3) = 0.2313 > 0 (+ve)

.. f(2) & f(3) have opposite signs

.. Root lies b/w 2 & 3

$$x_n = \frac{af(b) - bf(a)}{f(b) - f(a)}$$

N	a (-ve)	b (+ve)	f(a)	f(b)	$x_n = \frac{af(b) - bf(a)}{f(b) - f(a)}$	f(xn)
1	2	3	-0.5979	0.2313	2.7210	-0.0171
2	2.7210	3	-0.0171	0.2313	2.7402	-0.0003
3	2.7402	3	-0.0003	0.2313	2.7405	-0.0001
4	2.7405	3	-0.0001	0.2313	2.7406	-0.0004
5	2.7406	3	-0.0004	0.2313	2.7406	

.. x4 = x5

.. The root is 2.7406

3. Find root of equation x*e^x = 2

let f(x) = x*e^x - 2

put x=0 => f(0) = -2 (-ve)

put x=1 => f(1) = 1e - 2 = 0.7183 (+ve)

.. f(0) & f(1) have opposite signs

.. Root lies b/w 0 & 1

$$x_n = \frac{af(b) - bf(a)}{f(b) - f(a)}$$

N	a (-ve)	b (+ve)	f(a)	f(b)	$x_n = \frac{af(b) - bf(a)}{f(b) - f(a)}$	f(xn)
1	0	1	-2	0.7183	0.7358	-0.4643
2	0.7358	1	-0.4643	0.7183	0.8395	-0.0564
3	0.8395	1	-0.0564	0.7183	0.8512	-0.0061
4	0.8512	1	-0.0061	0.7183	0.8526	-0.0009
5	0.8526	1	-0.0009	0.7183	0.8526	-0.0002

For upto 3 decimals x4 = x5

.. The root is 0.852

The Iteration Method (or) The method of successive approximation method:

Suppose we want to find the root of the equation f(x)=0 reduce the equation as x=ø(x).

Suppose the initial approximation is x0 then the first approximation is x1=ø(x0)

The remaining successive approximation are x2=ø(x1), x3=ø(x2), ..., xn=ø(xn-1)

At one stage of the two successive approximation are equal say xn = xn-1 then the root is xn.

But the iteration method is applied when ø(x) is continuous i.e |ø'(x)| < 1

Iteration (Fixed Point) Method

The **Iteration Method** (also called the **Fixed Point Iteration Method**) is a numerical method used to find the approximate root of an equation by repeatedly substituting an initial approximation into a suitable function.

Advantages of Iteration (Fixed-Point Iteration) Method

1. It is **simple and easy to implement**.
2. It **does not require the derivative** of the function.
3. It requires **less computational effort** per iteration.

4. It is suitable for solving **nonlinear equations** when an appropriate iteration function is chosen.
5. The method can converge **rapidly** if the iteration function satisfies the convergence condition (i.e., $|\ g'(x)\ |< 1$near the root).

Disadvantages of Iteration (Fixed-Point Iteration) Method

1. **Convergence is not guaranteed** for every iteration function.
2. Choosing a suitable iteration function $g(x)$can be **difficult**.
3. The method may **converge slowly** compared to methods like the Newton–Raphson Method.
4. If the convergence condition is not satisfied, the method may **diverge or oscillate**.
5. The number of iterations required depends **strongly on the initial guess** and the choice of the iteration function.

Procedure (Algorithm):

Given the equation

$$f(x) = 0,$$

rewrite it in the form

$$x = \phi(x).$$

Choose an initial approximation x_0.
Calculate

$$x_1 = \phi(x_0),$$
$$x_2 = \phi(x_1),$$
$$x_3 = \phi(x_2),$$

and so on until two successive values are equal (or nearly equal).
The common value is the required root.

Condition for Convergence

The iteration method converges only if $|\ \phi'(x)\ |< 1$
in the interval containing the root.
If

$$|\ \phi'(x)\ |> 1,$$

the method diverges.

Problems

1. Find the root of x^3-x-11=0 using iteration method

let f(x) = x^3-x-11
put x=0 => f(0) = -11 < 0
put x=1 => f(1) = -11 < 0
put x=2 => f(2) = -5 < 0
put x=3 => f(3) = 13 > 0
.. f(2) & f(3) have opposite signs
.. Root lies b/w 2 & 3
x0 = (a+b)/2 = (2+3)/2 = 2.5
From given x^3-x-11=0
x^3 = x+11
x = (x+11)^(1/3)
let ø(x) = (x+11)^(1/3)
ø'(x) = 1/3*(x+11)^(-2/3) = 1/3*(x+11)^(-2/3)
|ø'(x)| < 1/3*(1)^(-2/3) < 1
.. |ø'(x)| < 1

$$x = \phi(x) = (x+11)^{1/3}$$

Take $x_0 = 2.5$

Iterations

$$x_1 = (2.5+11)^{1/3} = 2.3781$$
$$x_2 = (2.3781+11)^{1/3} = 2.3740$$
$$x_3 = (2.3740+11)^{1/3} = 2.3736$$
$$x_4 = (2.3736+11)^{1/3} = 2.3736$$

Since

$$x_3 = x_4,$$

the required root is

$$\boxed{x = 2.3736}$$

Problem 2

Solve $3x = \cos x + 1$ using the Iteration Method.

Solution:

Step 1: Write the equation $3x = \cos x + 1$

or

$$f(x) = \cos x - 3x + 1.$$

Step 2: Locate the root

Calculate

$$f(0) = \cos 0 - 3(0) + 1 = 2 > 0$$
$$f\left(\frac{\pi}{4}\right) = 0.7071 - 2.3562 + 1 = -0.6491 < 0$$
$$f\left(\frac{\pi}{2}\right) = 0 - 4.7124 + 1 = -3.7124 < 0$$

Since

$$f(0) > 0, f\left(\frac{\pi}{4}\right) < 0,$$

the root lies between

$$\boxed{\left(0, \frac{\pi}{4}\right)}$$

(or more generally within $\left(0, \frac{\pi}{2}\right)$).

Step 3: Convert into Iteration Form

$$3x = \cos x + 1$$
$$x = \frac{\cos x + 1}{3}$$

Therefore,

$$\boxed{\phi(x) = \frac{\cos x + 1}{3}}$$

Step 4: Check the Convergence Condition

Differentiate

$$\phi(x) = \frac{\cos x + 1}{3}$$
$$\phi'(x) = \frac{-\sin x}{3}$$

At

$$x = 0,$$
$$|\phi'(0)| = 0 < 1$$

At

$$x = \frac{\pi}{2},$$
$$|\phi'\left(\frac{\pi}{2}\right)| = \frac{1}{3} = 0.333 < 1$$

Hence,

$$|\phi'(x)| < 1.$$

Therefore, the Iteration Method is applicable.

Step 5: Iteration Table

Take

$$x_0 = 0$$

and use

$$x_{n+1} = \frac{\cos x_n + 1}{3}.$$

Iteration	Formula	Value
x_0	Initial value	0.00000
x_1	$\frac{\cos 0 + 1}{3}$	0.66667
x_2	$\frac{\cos 0.66667 + 1}{3}$	0.59530
x_3	$\frac{\cos 0.59530 + 1}{3}$	0.60933
x_4	$\frac{\cos 0.60933 + 1}{3}$	0.60668
x_5	$\frac{\cos 0.60668 + 1}{3}$	0.60718
x_6	$\frac{\cos 0.60718 + 1}{3}$	0.60709
x_7	$\frac{\cos 0.60709 + 1}{3}$	0.60710
x_8	$\frac{\cos 0.60710 + 1}{3}$	0.60710

Since

$$x_7 = x_8,$$

the iterations have converged.

Final Answer

$$\boxed{x = 0.60710}$$

Exam Steps (5 Marks)

1. Write the equation $f(x) = 0$.
2. Rewrite it as $x = \phi(x)$.
3. Check the convergence condition $| \phi'(x) | < 1$.
4. Choose an initial approximation x_0.
5. Compute $x_1, x_2, x_3, \ldots$.
6. Stop when two successive values are equal.
7. Write the final root.

---

2 MARKS

1. What is the Iteration Method?

It is a numerical method in which the equation is written as $x = \phi(x)$, and successive approximations are used to obtain the root.

2. What is the convergence condition?

$$\boxed{| \phi'(x) | < 1}$$

3. Formula used in the Iteration Method

$$\boxed{x_{n+1} = \phi(x_n)}$$

Newton–Raphson Method

The Newton–Raphson Method is one of the fastest numerical methods used to find the approximate root of an equation.
It starts with an initial approximation and improves it by repeated iterations.

Formula

If

$$\boldsymbol{f(x) = 0},$$

then the Newton–Raphson iteration formula is

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

where

- x_n= current approximation
- x_{n+1}= next approximation
- $f'(x)$= derivative of $f(x)$

The iterations are continued until two successive values become equal (or nearly equal).

Algorithm (Procedure)

Step 1

Write the equation in the form

$$f(x) = 0.$$

Step 2

Find the derivative

$$f'(x).$$

Step 3

Choose an initial approximation x_0.

Step 4

Use the formula

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

Step 5

Repeat the process until the desired accuracy is obtained.

Advantages

- Very fast convergence.
- Requires fewer iterations.
- Highly accurate.

Disadvantages

- Requires the derivative $f'(x)$.

- **A poor initial guess may lead to divergence.**
- **Cannot be applied if $f'(x) = 0$.**

Solved Problem 1

Find the root of

$$x^3 - x - 1 = 0$$

using the Newton–Raphson Method.

Solution:

Step 1: Write the function

$$f(x) = x^3 - x - 1.$$

Step 2: Locate the root

Calculate

$$f(1) = 1 - 1 - 1 = -1 < 0$$
$$f(2) = 8 - 2 - 1 = 5 > 0$$

Since the signs are opposite,
the root lies between

1 and 2.

Step 3: Initial Approximation

Take

$$x_0 = \frac{1+2}{2} = 1.5.$$

Step 4: Differentiate

$$f'(x) = 3x^2 - 1.$$

Step 5: Apply Newton–Raphson Formula

$$x_{n+1} = x_n - \frac{x_n^3 - x_n - 1}{3x_n^2 - 1}.$$

Iteration Table

Iteration (n)	x_n	x_{n+1}
0	1.5000	1.3478

Iteration (n)	x_n	x_{n+1}
1	1.3478	1.3252
2	1.3252	1.3247
3	1.3247	1.3247

Since

$$x_3 = x_4,$$

the iterations have converged.

Final Answer

$$\boxed{x = 1.3247}$$

Solved Problem 2 (Partial)

Find the root of

$$e^x \sin x = 1$$

using the Newton–Raphson Method.

Solution:

Step 1: Write the function

$$f(x) = e^x \sin x - 1.$$

Step 2: Test the function

At

$$x = 0,$$
$$f(0) = e^0 \sin 0 - 1 = 1 \times 0 - 1 = -1.$$

Since

$$f(0) < 0,$$

we test another value (such as $x = 1$) to locate the interval containing the root.

Step 3: Differentiate

Using the product rule,

$$\frac{d}{dx}(e^x \sin x) = e^x \sin x + e^x \cos x.$$

Therefore,

$$\boxed{f'(x) = e^x(\sin x + \cos x)}$$

Step 4: Newton–Raphson Formula

The iteration formula becomes

$$\boxed{x_{n+1} = x_n - \frac{e^{x_n}\sin x_n - 1}{e^{x_n}(\sin x_n + \cos x_n)}}$$

After choosing a suitable initial approximation, this formula is used repeatedly until two successive values are equal.

Formula Summary

Newton–Raphson Formula

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

For $x^3 - x - 1 = 0$

$$\boxed{x_{n+1} = x_n - \frac{x_n^3 - x_n - 1}{3x_n^2 - 1}}$$

For $e^x \sin x = 1$

$$\boxed{x_{n+1} = x_n - \frac{e^{x_n}\sin x_n - 1}{e^{x_n}(\sin x_n + \cos x_n)}}$$

Exam Points

1. Newton–Raphson Method:

A numerical method used to find the root of an equation by successive approximations.

2. Main Formula:

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

3. Advantage:

Converges very rapidly (faster than the Bisection Method).

4. Disadvantage:

Requires the derivative of the function and may fail if $f'(x) = 0$.

5. Root of $x^3 - x - 1 = 0$:

$$\boxed{x \approx 1.3247}$$

Note: Compared with the Bisection Method (which gave approximately 1.322 after several iterations), the Newton–Raphson Method reaches the more accurate value 1.3247 in only a few iterations because it converges much faster.

---

3. Using Newton–Raphson Method find the root of

$$x^4 - x - 9 = 0$$

Solution:

Let

$$f(x) = x^4 - x - 9$$

Then, $f'(x) = 4x^3 - 1$

Put $x = 0$, $f(0) = 0 - 0 - 9 = -9$ **(-ve)**

Put $x = 1$, $f(1) = 1 - 1 - 9 = -9$ **(-ve)**

Put $x = 2$, $f(2) = 16 - 2 - 9 = 5$ **(+ve)**

Since $f(1)$ **and** $f(2)$ **have opposite signs,**

$\therefore$ **Root lies between 1 and 2.**

Take the initial approximation,

$$x_0 = \frac{1+2}{2} = 1.5$$

By Newton–Raphson Method,

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

Iteration Table

n	x_n	x_{n+1}
0	1.5000	1.9230
1	1.9230	1.8245
2	1.8245	1.8137
3	1.8137	1.8136
4	1.8136	1.8136

Since

$$x_3 = x_4,$$

$$\boxed{\therefore \text{ The root is } 1.8136}$$

Interpolation

Let $y = f(x)$ be a function. Let

$$y_0, y_1, y_2, \dots, y_n$$

be the values of y corresponding to the values of x,

$$x_0, x_1, x_2, \dots, x_n.$$

These values are called entries.

The process of computing the value of the function $y = f(x)$ within the given range of data is called Interpolation.

The process of computing the value of the function outside the given range is called Extrapolation.

To determine the value of the function $f(x)$ at some intermediate value of x, the following difference operators are used:

1. Forward Difference Operator (Δ)
2. Backward Difference Operator (∇)
3. Central Difference Operator (δ)

---

1. Forward Difference

Consider the function

$$y = f(x),$$

where

$$y_0, y_1, \dots, y_n$$

are the values corresponding to

$$x_0, x_1, \dots, x_n.$$

First Order Forward Difference

$$\begin{aligned}
\Delta y_0 &= y_1 - y_0 \\
\Delta y_1 &= y_2 - y_1 \\
\Delta y_2 &= y_3 - y_2 \\
\Delta y_3 &= y_4 - y_3 \\
&\dots \\
\Delta y_{n-1} &= y_n - y_{n-1}
\end{aligned}$$

Second Order Forward Difference

$$\begin{aligned}
&\Delta^2 y_0 = \Delta y_1 - \Delta y_0\\
&\Delta^2 y_1 = \Delta y_2 - \Delta y_1\\
&\Delta^2 y_2 = \Delta y_3 - \Delta y_2\\
&\dots\\
&\Delta^2 y_{n-2} = \Delta y_{n-1} - \Delta y_{n-2}
\end{aligned}$$

Third Order Forward Difference

$$\begin{aligned}
&\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0\\
&\Delta^3 y_1 = \Delta^2 y_2 - \Delta^2 y_1\\
&\dots\\
&\Delta^3 y_{n-3} = \Delta^2 y_{n-2} - \Delta^2 y_{n-3}
\end{aligned}$$

General r^{th} Order Forward Difference

$$\boxed{\Delta^r y_i = \Delta^{r-1} y_{i+1} - \Delta^{r-1} y_i}$$

where

$$r = 1, 2, \dots, n.$$

Forward Difference Table

x	y	Δy	Δ²y	Δ³y	Δ⁴y
x_0	y_0	$\Delta y_0 = y_1 - y_0$	$\Delta^2 y_0 = \Delta y_1 - \Delta y_0$	$\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0$	$\Delta^4 y_0 = \Delta^3 y_1 - \Delta^3 y_0$
x_1	y_1	$\Delta y_1 = y_2 - y_1$	$\Delta^2 y_1 = \Delta y_2 - \Delta y_1$	$\Delta^3 y_1 = \Delta^2 y_2 - \Delta^2 y_1$	
x_2	y_2	$\Delta y_2 = y_3 - y_2$	$\Delta^2 y_2 = \Delta y_3 - \Delta y_2$		
x_3	y_3	$\Delta y_3 = y_4 - y_3$			
x_4	y_4				

Example

Q1. Write down the Forward Difference Table for the data:

X	4	6	8	10
Y	1	3	8	16

Solution:

Forward Difference Table

For the data:

X	y	Δy	$\Delta^2 y$	$\Delta^3 y$
4	1	$\Delta y_0 = 2$	$\Delta^2 y_0 = 3$	$\Delta^3 y_0 = 0$
6	3	$\Delta y_1 = 5$	$\Delta^2 y_1 = 3$	
8	8	$\Delta y_2 = 8$		
10	16			

Newton's Forward Difference Formula (Newton's Forward Interpolation Formula)

Let

$$y = f(x)$$

where

$$x_0, x_1, x_2, \dots, x_n$$

are equally spaced values of x, and

$$y_0, y_1, y_2, \dots, y_n$$

are the corresponding values of y.
Suppose it is required to find the value of the function at some point x, where $x$lies between $x_0$and x_n, and the value of $x$is near the beginning of the interval. Then we use Newton's Forward Interpolation Formula.

Formula

$$y = y_0 + P\Delta y_0 + \frac{P(P-1)}{2!}\Delta^2 y_0 + \frac{P(P-1)(P-2)}{3!}\Delta^3 y_0 + \cdots + \frac{P(P-1)(P-2)\cdots(P-n+1)}{n!}\Delta^n y_0$$

where

$$\boxed{P = \frac{x - x_0}{h}}$$

Here,

- x= value at which interpolation is required
- x_0= initial value of x
- h= common interval (equal spacing)

This formula is also called Newton's Gregory Forward Interpolation Formula.

Example

Find $f(1.6)$using Newton's Forward Interpolation Formula for the following data:

x	1	1.4	1.8	2.2
y	3.49	4.82	5.96	6.50

Forward Difference Table

x	Y	Δy	Δ^2y	Δ^3y
1.0	3.49			
		Δy_0 = 1.33		
1.4	4.82		$\Delta^2 y_0$ = −0.19	
		Δy_1 = 1.14		$\Delta^3 y_0$ = −0.41
1.8	5.96		$\Delta^2 y_1$ = −0.60	
		Δy_2 = 0.54		
2.2	6.50			

From the given data, the difference between successive values of $x$is equal.

$$h = 0.4$$

We have to find

$$f(1.6)$$

using Newton's Forward Interpolation Formula.

Using

$$y = y_0 + P\Delta y_0 + \frac{P(P-1)}{2!}\Delta^2 y_0 + \frac{P(P-1)(P-2)}{3!}\Delta^3 y_0 + \cdots$$

where

$$P = \frac{x - x_0}{h}$$

Here,

- $x = 1.6$
- $x_0 = 1$
- $h = 0.4$

Therefore,

$$P = \frac{1.6 - 1}{0.4} = \frac{0.6}{0.4} = 1.5$$

Using Newton's Forward Interpolation Formula,

$$y = 3.49 + 1.5(1.33) + \frac{1.5(1.5-1)}{2!}(-0.19) + \frac{1.5(0.5)(-0.5)}{3!}(-0.41)$$

$$= 3.49 + 1.995 - 0.07125 + 0.025625$$

$$\boxed{y = 5.439375}$$

Hence,

$$\boxed{f(1.6) = 5.439375}$$

Example

Find $f(2.5)$ using Newton's Forward Interpolation Formula for the following data.

Given Data

x	0	1	2	3	4	5	6
y	0	1	16	81	256	625	1296

From the given data,

$$h = 1$$

We have to find

$$f(2.5)$$

using Newton's Forward Interpolation Formula.

Forward Difference Table

x	Y	Δy	Δ²y	Δ³y	Δ⁴y	Δ⁵y	Δ⁶y
0	0						
1	1	1	14	36	24	0	0
2	16	15	50	60	24	0	
3	81	65	110	84	24		
4	256	175	194	108			
5	625	369	302				
6	1296	671					

Using Newton's Forward Formula,

$$y = y_0 + P\Delta y_0 + \frac{P(P-1)}{2!}\Delta^2 y_0 + \frac{P(P-1)(P-2)}{3!}\Delta^3 y_0 + \frac{P(P-1)(P-2)(P-3)}{4!}\Delta^4 y_0 + \frac{P(P-1)(P-2)(P-3)(P-4)}{5!}\Delta^5 y_0 + \frac{P(P-1)(P-2)(P-3)(P-4)(P-5)}{6!}\Delta^6 y_0$$

where

$$P = \frac{2.5-0}{1} = 2.5$$

Substituting,

$$y = 0 + 0.5 + 26.25 + 11.25 + 0.9375$$
$$\boxed{y = 39.0625}$$

Hence,

$$\boxed{f(2.5) = 39.0625}$$

Example

For $x = 0,1,2,3,4$ and $f(x) = 1,14,15,5,6,$

find the interpolation polynomial using Newton's Forward Interpolation Formula.

Solution:

From the given data,

$$h = 1, x_0 = 0$$
$$P = \frac{x - x_0}{h} = x$$

Difference Table

x	Y	Δy	Δ²y	Δ³y	Δ⁴y
0	1				
1	14	13	-12		
2	15	1	-11	1	21
3	5	-10	11	22	
4	6	1			

Using Newton's Forward Formula,

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

Substituting,

$$= 1 + x(13) + \frac{x(x-1)}{2}(-12) + \frac{x(x-1)(x-2)}{6}(1)$$

Expanding,

$$= 1 + 13x - 6x(x-1) + \frac{x(x-1)(x-2)}{6}$$
$$= \frac{x^3 - 39x^2 + 116x + 6}{6}$$

Therefore,

$$\boxed{f(x) = \frac{1}{6}x^3 - \frac{13}{2}x^2 + \frac{58}{3}x + 1}$$

Example

If $u_0 = 1,\ u_1 = 0,\ u_2 = 5,\ u_3 = 22,\ u_4 = 57,$ find $u(0.5)$.

Solution:

Difference Table

u	y	Δy	Δ²y	Δ³y	Δ⁴y
0	1	-1	6	6	0
1	0	5	12	6	
2	5	17	18		
3	22	35			
4	57				

Given,

$$x = 0.5, x_0 = 0, h = 1$$

Hence,

$$P = \frac{0.5 - 0}{1} = 0.5$$

Using Newton's Forward Formula,

$$y = y_0 + P\Delta y_0 + \frac{P(P-1)}{2!}\Delta^2 y_0 + \frac{P(P-1)(P-2)}{3!}\Delta^3 y_0$$

Substituting,

$$= 1 + 0.5(-1) + \frac{0.5(-0.5)}{2}(6) + \frac{0.5(-0.5)(-1.5)}{6}(6)$$
$$= 1 - 0.5 - 0.75 + 0.375$$
$$\boxed{y = 0.125}$$

Therefore,

$$\boxed{u(0.5) = 0.125}$$

Backward Differences

Let

$$y = f(x)$$

be the function.
Let

$$y_0, y_1, \dots, y_n$$

be the functional values corresponding to

$$x_0, x_1, \dots, x_n.$$

Then the backward differences are

$$\begin{aligned} \nabla y_1 &= y_1 - y_0, \\ \nabla y_2 &= y_2 - y_1, \\ \nabla y_3 &= y_3 - y_2, \\ &\dots \\ \nabla y_n &= y_n - y_{n-1}. \end{aligned}$$

These are called **first-order backward differences.**
The **second-order backward differences** are

$$\begin{aligned} \nabla^2 y_2 &= \nabla y_2 - \nabla y_1, \\ \nabla^2 y_3 &= \nabla y_3 - \nabla y_2, \\ &\dots \\ \nabla^2 y_n &= \nabla y_n - \nabla y_{n-1}. \end{aligned}$$

In general,

$$\boxed{\nabla^r y_n = \nabla^{r-1} y_n - \nabla^{r-1} y_{n-1}}$$

where

$$n = 1,2,3, \dots$$

These are called the r^{th}**-order backward differences**.

Backward Difference Table

x	y	∇y	∇²y	∇³y	∇⁴y
x_0	y_0				
x_1	y_1	$\nabla y_1 = y_1 - y_0$			
x_2	y_2	$\nabla y_2 = y_2 - y_1$	$\nabla^2 y_2 = \nabla y_2 - \nabla y_1$		
x_3	y_3	$\nabla y_3 = y_3 - y_2$	$\nabla^2 y_3 = \nabla y_3 - \nabla y_2$	$\nabla^3 y_3 = \nabla^2 y_3 - \nabla^2 y_2$	
x_4	y_4	$\nabla y_4 = y_4 - y_3$	$\nabla^2 y_4 = \nabla y_4 - \nabla y_3$	$\nabla^3 y_4 = \nabla^2 y_4 - \nabla^2 y_3$	$\nabla^4 y_4 = \nabla^3 y_4 - \nabla^3 y_3$

Newton's Backward Difference Formula (Newton's Backward Interpolation Formula)

Let the function

$$y = f(x)$$

satisfy the values

$$y_0, y_1, \dots, y_n$$

corresponding to

$$x_0, x_1, \dots, x_n$$

of the independent variable x.
Let the interval between successive values of $x$be equal.
Suppose it is required to find the value of $y$at some point $x$lying inside the interval (x_0, x_n), and the required value is **near the end of the table**.
Then Newton's Backward Difference Formula is used.

Formula

$$\boxed{y = f(x) = y_n + P\nabla y_n + \frac{P(P+1)}{2!}\nabla^2 y_n + \frac{P(P+1)(P+2)}{3!}\nabla^3 y_n + \cdots}$$

where

$$\boxed{P = \frac{x - x_n}{h}}$$

Here,

- x_n= last value of x
- x= value where interpolation is required
- h= common interval

Notes

1. Newton's Backward Difference Formula is applicable only for **equal intervals**.
2. It is useful when the required value lies **near the end of the table**.

Example 1

Find the value of y at $x = 9$ using Newton's Backward Difference Formula.
Given Data

x	2	5	8	11
y	94.8	87.9	81.3	75.1

Solution:

Backward Difference Table

x	Y	∇y	∇²y	∇³y
2	94.8			
5	87.9	-6.9	0.3	
8	81.3	-6.6	0.4	0.1
11	75.1	-6.2		

Here,

$$x = 9, x_n = 11, h = 3$$

Therefore,

$$P = \frac{9-11}{3} = -\frac{2}{3} \approx -0.67$$

Using Newton's Backward Formula,

$$y = y_n + P\nabla y_n + \frac{P(P+1)}{2!}\nabla^2 y_n + \frac{P(P+1)(P+2)}{3!}\nabla^3 y_n$$

Substituting,

$$\begin{aligned} y &= 75.1 + (-0.67)(-6.2) \\ &+ \frac{(-0.67)(0.33)}{2}(0.4) \\ &+ \frac{(-0.67)(0.33)(1.33)}{6}(0.1) \\ &= 75.1 + 4.154 - 0.04422 - 0.0049105 \end{aligned}$$

$$\boxed{y = 79.204}$$

Example 2

Use Newton's Backward Difference Formula to find $f(7.5)$.

Given Data

x	1	2	3	4	5	6	7	8
y	1	8	27	64	125	216	343	512

Solution:

Difference Table

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$	$\nabla^4 y$	$\nabla^5 y$
1	1					
2	8	7				
3	27	19	12			
4	64	37	18	6		
5	125	61	24	6	0	
6	216	91	30	6	0	0
7	343	127	36	6	0	0
8	512	169	42	6	0	0

Here,

$$x = 7.5, x_n = 8, h = 1$$

$$P = \frac{7.5 - 8}{1} = -0.5$$

Using Newton's Backward Formula,

$$\begin{aligned} y &= 512 + (-0.5)(169) \\ &+ \frac{(-0.5)(0.5)}{2}(42) \\ &+ \frac{(-0.5)(0.5)(1.5)}{6}(6) \\ &= 512 - 84.5 - 5.25 - 0.375 \\ &\boxed{f(7.5) = 421.875} \end{aligned}$$

Example 3

Using Backward Difference Formula, find the polynomial passing through

$$(3,6),\ (4,24),\ (5,60),\ (6,120)$$

Solution:

Difference Table

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$
3	6	18 36 60	18 24	6
4	24			

x	y	∇y	∇²y	∇³y
5	60			
6	120			

Here,

$$x_n = 6, h = 1, P = x - 6$$

Using Newton's Backward Formula,

$$y = 120 + P(60) + \frac{P(P+1)}{2}(24) + \frac{P(P+1)(P+2)}{6}(6)$$

Substituting $P = x - 6$,

$$y = 120 + 60(x-6) + 12(x-6)(x-5) + (x-6)(x-5)(x-4)$$

Expanding,

$$\boxed{y = x^3 - 3x^2 + 2x}$$

Interpolation with Unequal Intervals (Lagrange's Interpolation)

Let

$$y = f(x)$$

be the function, and let

$$f(x_0),\ f(x_1),\ f(x_2),\ \dots,\ f(x_n)$$

be the values of $y$corresponding to

$$x_0,\ x_1,\ x_2,\ \dots,\ x_n.$$

If the values of $x$are **not equally spaced**, then Newton's interpolation formulas cannot be used.
In such cases, **Lagrange's Interpolation Formula** is used.
Lagrange's Formula

$$\boxed{f(x) = \frac{(x-x_1)(x-x_2)\cdots(x-x_n)}{(x_0-x_1)(x_0-x_2)\cdots(x_0-x_n)}f(x_0) + \frac{(x-x_0)(x-x_2)\cdots(x-x_n)}{(x_1-x_0)(x_1-x_2)\cdots(x_1-x_n)}f(x_1) + \cdots}$$

Lagrange Interpolation Formula

$$y = f(x) = \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)}f(x_0) + \frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)}f(x_1)$$
$$+ \frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}f(x_2) + \cdots + \frac{(x-x_0)(x-x_1)\cdots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\cdots(x_n-x_{n-1})}f(x_n)$$

Example 1

Using the Lagrange Interpolation Formula find the value of $y(10)$from the following table.

x	**5**	**6**	**9**	**11**
y	12	13	14	16

Solution:

From the given table, the difference between the x-values is **unequal**, so we use the **Lagrange Interpolation Formula (LIF)**.

Given

$$x_0 = 5, x_1 = 6, x_2 = 9, x_3 = 11$$
$$f(x_0) = 12, f(x_1) = 13, f(x_2) = 14, f(x_3) = 16$$

Using Lagrange interpolation,

$$y = f(10) = \frac{(10-6)(10-9)(10-11)}{(5-6)(5-9)(5-11)}(12) + \frac{(10-5)(10-9)(10-11)}{(6-5)(6-9)(6-11)}(13) + \frac{(10-5)(10-6)(10-11)}{(9-5)(9-6)(9-11)}(14) + \frac{(10-5)(10-6)(10-9)}{(11-5)(11-6)(11-9)}(16)$$

Substituting,

$$= 2 + (-4.33) + 11.66 + 5.33$$
$$\boxed{y(10) \approx 14.66 = \frac{44}{3}}$$

Example 2

Calculate $f(10)$ given the following values:

x	**1**	**7**	**15**
$f(x)$	168	192	336

Solution:

Given

$$x_0 = 1, x_1 = 7, x_2 = 15$$
$$f(x_0) = 168, f(x_1) = 192, f(x_2) = 336$$

Since the x-values are **unequally spaced**, use the **Lagrange Interpolation Formula**.

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

For $x = 10$,

$$y = f(10) = \frac{(10 - 7)(10 - 15)}{(1 - 7)(1 - 15)}(168) + \frac{(10 - 1)(10 - 15)}{(7 - 1)(7 - 15)}(192) + \frac{(10 - 1)(10 - 7)}{(15 - 1)(15 - 7)}(336)$$

Simplifying,

$$= \frac{3(-5)(168)}{(-6)(-14)} + \frac{9(-5)(192)}{6(-8)} + \frac{9(3)(336)}{14(8)}$$
$$= -30 + 180 + 126$$
$$\boxed{f(10) = 276}$$

PRACTICE QUESTIONS:

1.Find the root of $x^3 + x - 1 = 0$ using Newton Raphson method
2.Find the root of $x^3 - 9 = 0$ using Newton Raphson method
3.Solve $x^3 - x - 2 = 0$ using the Iteration method.
4. Solve $x^3 - 4x - 9 = 0$by the Bisection method.
5. Find the root of $x^2 - 5 = 0$using the Bisection method.
6. Solve $e^x - 3x = 0$using the Bisection method.
7. Solve $x^3 - 2x - 5 = 0$using the Regula-Falsi method
8. Find the root of $x\log_{10} x - 1.2 = 0$using the Regula-Falsi method.
9. Solve $e^{-x} - x = 0$using the Newton-Raphson method.
10. Find the positive root of $x - \sin x - 0.5 = 0$using the Newton-Raphson method.

Unit–2

Numerical Differentiation, Numerical Integration and Curve Fitting

Unit–2

Numerical Differentiation, Numerical Integration and Curve Fitting

Numerical Integration

Numerical integration is the approximate computation of a definite integral

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

using numerical techniques. Numerical integration is also known as **Numerical Quadrature**. The most straightforward numerical integration technique uses the **Newton–Cotes formula**.

Newton–Cotes Formula

Let $y_0, y_1, y_2, \dots, y_n$ be the values of

$$y = f(x)$$

corresponding to

$$a = x_0,\ x_1,\ x_2,\ \dots,\ x_n = b$$

which are equally spaced with step size h.

Then,

$$x_r = x_0 + rh, x_n = x_0 + nh$$

By Newton's Forward Interpolation Formula,

$$y = y_0 + \frac{P}{1!}\Delta y_0 + \frac{P(P-1)}{2!}\Delta^2 y_0 + \frac{P(P-1)(P-2)}{3!}\Delta^3 y_0 + \cdots \qquad (1)$$

where

$$P = \frac{x - x_0}{h}$$

Now,

$$\int_{x_0}^{x_n} y\,dx = \int_{x_0}^{x_n} \left[y_0 + \frac{P}{1!}\Delta y_0 + \frac{P(P-1)}{2!}\Delta^2 y_0 + \frac{P(P-1)(P-2)}{3!}\Delta^3 y_0 + \cdots \right] dx$$

where

- h= interval size (step size)
- n= number of sub-intervals
- $a$and b= limits of integration

Since

$$x = x_0 + Ph,$$

therefore,

$$dx = h\,dP.$$

Hence,

$$\int_{x_0}^{x_n} y\,dx = h\int_0^n \left[y_0 + P\Delta y_0 + \frac{P^2 - P}{2}\Delta^2 y_0 + \frac{P^3 - 3P^2 + 2P}{6}\Delta^3 y_0 + \cdots\right] dP$$

Integrating,

$$\int_{x_0}^{x_n} y\,dx = h\left[ny_0 + \frac{n^2}{2}\Delta y_0 + \frac{1}{2}\left(\frac{n^3}{3} - \frac{n^2}{2}\right)\Delta^2 y_0 + \cdots\right]$$

This is called the **Newton–Cotes Quadrature Formula**.

Trapezoidal Rule

Put $n = 1$in the above formula.
Then,

$$\int_{x_0}^{x_1} y\,dx = h\left(y_0 + \frac{1}{2}\Delta y_0\right)$$

Since higher-order differences do not exist for $n = 1$,

$$\Delta y_0 = y_1 - y_0.$$

Therefore,

$$\int_{x_0}^{x_1} y\,dx = h\left[y_0 + \frac{1}{2}(y_1 - y_0)\right]$$

or,

$$\boxed{\int_{x_0}^{x_1} y\,dx = \frac{h}{2}(y_0 + y_1)}$$

This is known as the **Trapezoidal Rule**.
Putting $n = 2$,

$$\int_{x_0}^{x_2} y\,dx = h\left[y_0+\Delta y_0+\frac{1}{3}\Delta^2 y_0\right]$$

which simplifies to

$$\int_{x_0}^{x_2} y\,dx = \frac{h}{2}(y_0+y_2)$$

Similarly,

$$\int_{x_0+(n-1)h}^{x_0+nh} y\,dx = \frac{h}{2}(y_{n-1}+y_n)$$

Since $y = f(x)$,

$$\int_{x_0}^{x_n} f(x)\,dx = \int_{x_0}^{x_1} f(x)\,dx + \int_{x_1}^{x_2} f(x)\,dx + \cdots + \int_{x_{n-1}}^{x_n} f(x)\,dx$$

Therefore,

$$\begin{aligned}\int_{x_0}^{x_n} f(x)\,dx &= \frac{h}{2}(y_0+y_1)+\frac{h}{2}(y_1+y_2)+\cdots+\frac{h}{2}(y_{n-1}+y_n)\\ &= \frac{h}{2}[(y_0+y_n)+2(y_1+y_2+\cdots+y_{n-1})]\end{aligned}$$

Hence,

$$\boxed{\int_{x_0}^{x_n} f(x)\,dx = \frac{h}{2}[(\text{Sum of first and last ordinates}) + 2\times(\text{Sum of remaining ordinates})]}$$

Simpson's $\frac{1}{3}$Rule:

This is another popular numerical integration method.
By putting $n = 2$in the Newton–Cotes Quadrature Formula,

$$\boxed{\int_{x_0}^{x_n} f(x)\,dx = \frac{h}{3}[(y_0+y_n)+4(y_1+y_3+\cdots+y_{n-1})+2(y_2+y_4+\cdots+y_{n-2})]}$$

where

$$h = \frac{x_n - x_0}{n} = \frac{b-a}{n}$$

or

$$\boxed{\frac{h}{3}[(\text{Sum of first and last ordinates}) + 4 \times (\text{Sum of odd ordinates}) + 2 \times (\text{Sum of even ordinates})]}$$

This is known as **Simpson's $\frac{1}{3}$Rule**.

Simpson's $\frac{3}{8}$Rule:

$$\boxed{\int_{x_0}^{x_n} f(x)\,dx = \frac{3h}{8}[(y_0 + y_n) + 3(y_1 + y_2 + y_4 + y_5 + y_7 + y_8 + \cdots) + 2(y_3 + y_6 + y_9 + \cdots)]}$$

Example 1

Evaluate

$$\int_0^1 x^3\ dx$$

using the **Trapezoidal Rule** with **5 subintervals** and compare with the exact value.

Given

$$a = 0, b = 1, n = 5$$
$$h = \frac{b-a}{n} = \frac{1-0}{5} = 0.2$$
$$f(x) = x^3$$

Table

x	0	0.2	0.4	0.6	0.8	1
$y = f(x) = x^3$	0	0.008	0.064	0.216	0.512	1

Using the Trapezoidal Rule,

$$\int_0^1 x^3\ dx = \frac{h}{2}[(y_0 + y_5) + 2(y_1 + y_2 + y_3 + y_4)]$$
$$= \frac{0.2}{2}[(0+1) + 2(0.008 + 0.064 + 0.216 + 0.512)]$$
$$= 0.1[1 + 2(0.800)]$$
$$= 0.1(2.600) = 0.2600$$

Exact Value

$$\int_0^1 x^3 \, dx = \left[\frac{x^4}{4}\right]_0^1 = \frac{1}{4} = 0.2500$$

Comparison

- Trapezoidal Rule = **0.2600**
- Exact Value = **0.2500**

Example 2

Use the Trapezoidal Rule with $n = 4$to estimate

$$\int_0^1 \frac{1}{1+x^2} \, dx$$

Solution:

Given

$$a = 0, b = 1, n = 4$$
$$h = \frac{1-0}{4} = 0.25$$
$$f(x) = \frac{1}{1+x^2}$$

Table

x	**0**	**0.25**	**0.50**	**0.75**	**1**
$y = f(x)$	1	0.9412	0.8000	0.6400	0.5000

Using the Trapezoidal Rule,

$$\int_0^1 \frac{1}{1+x^2} \, dx = \frac{0.25}{2}[(1+0.5) + 2(0.9412 + 0.8 + 0.64)]$$
$$= 0.125[1.5 + 2(2.3812)]$$
$$= 0.125(6.2624)$$
$$\boxed{= 0.7828 \text{ (approximately)}}$$

Exercise

Evaluate

$$\int_0^{1.2} e^x \, dx$$

taking **6 intervals** using the **Trapezoidal Rule**.

Simpson's $\frac{1}{3}$Rule

$$\int_a^b f(x)\,dx = \frac{h}{3}[(y_0 + y_n) + 4(y_1 + y_3 + \cdots + y_{n-1}) + 2(y_2 + y_4 + \cdots + y_{n-2})]$$

where

$$h = \frac{b-a}{n}$$

Simpson's $\frac{3}{8}$Rule

$$\int_a^b f(x)\,dx = \frac{3h}{8}[(y_0 + y_n) + 2(y_3 + y_6 + \cdots) + 3(y_1 + y_2 + \cdots + y_{n-1})]$$

This method is applicable only when $n$is a multiple of **3**. It is not as accurate as Simpson's $\frac{1}{3}$rule.

Example

Given that

x	**4**	**4.2**	**4.4**	**4.6**	**4.8**	**5.0**	**5.2**
$y = \log_e x$	1.3863	1.4351	1.4816	1.5261	1.5686	1.6094	1.6487

Evaluate

$$\int_4^{5.2} \log_e x\,dx$$

using:

1. Trapezoidal Rule
2. Simpson's $\frac{1}{3}$Rule
3. Simpson's $\frac{3}{8}$Rule

Given

$$h = 0.2$$

(i) Trapezoidal Rule

$$\int_4^{5.2} \log_e x\,dx = \frac{0.2}{2}[(1.3863 + 1.6487) + 2(1.4351 + 1.4816 + 1.5261 + 1.5686 + 1.6094)]$$
$$= 0.1\,[3.0350 + 2(7.6208)]$$
$$= 0.1(18.2766)$$
$$\boxed{= 1.82766}$$

(ii) Simpson's $\frac{1}{3}$Rule

$$\int_4^{5.2} \log_e x\,dx = \frac{0.2}{3}[(1.3863 + 1.6487) + 4(1.4351 + 1.5261 + 1.6094) + 2(1.4816 + 1.5686)]$$
$$= 0.0667\,(3.0350 + 18.2824 + 6.1004)$$
$$\boxed{= 1.8288}$$

(iii) Simpson's $\frac{3}{8}$Rule

$$\int_4^{5.2} \log_e x\,dx = \frac{3(0.2)}{8}[(1.3863 + 1.6487) + 3(1.4351 + 1.4816 + 1.5686 + 1.6094) + 2(1.5261)]$$
$$= 0.0750\,(3.0350 + 18.2841 + 3.0522)$$
$$\boxed{= 1.8278}$$

Example

Divide the range into **10 equal parts**. Find the approximate value of

$$\int_0^{\pi} \sin x\,dx$$

using:

- Trapezoidal Rule
- Simpson's $\frac{1}{3}$Rule
- Simpson's $\frac{3}{8}$Rule

Given

$$a = 0, b = \pi, n = 10$$

$$h = \frac{b-a}{n} = \frac{\pi}{10}$$

Table of values

x	0	$\frac{\pi}{10}$	$\frac{2\pi}{10}$	$\frac{3\pi}{10}$	$\frac{4\pi}{10}$	$\frac{5\pi}{10}$	$\frac{6\pi}{10}$	$\frac{7\pi}{10}$	$\frac{8\pi}{10}$	$\frac{9\pi}{10}$	π
$y = \sin x$	0	0.3090	0.5878	0.8090	0.9511	1.0000	0.9511	0.8090	0.5878	0.3090	0

(i) Trapezoidal Rule

$$\int_0^{\pi} \sin x \, dx = \frac{\pi}{10(2)}[(0+0) + 2(0.3090 + 0.5878 + 0.8090 + 0.9511 + 1 + 0.9511 + 0.8090 + 0.5878 + 0.3090)]$$

$$= 0.1571(12.6276)$$

$$\boxed{= 1.9838}$$

(ii) Simpson's $\frac{1}{3}$Rule

$$\int_0^{\pi} \sin x \, dx = \frac{\pi}{30}[(0+0) + 4(0.3090 + 0.8090 + 1 + 0.8090 + 0.3090) + 2(0.5878 + 0.9511 + 0.9511 + 0.5878)]$$

$$= 0.1047(19.099)$$

$$\boxed{= 1.9997}$$

(iii) Simpson's $\frac{3}{8}$Rule

$$\int_0^{\pi} \sin x \, dx = \frac{3\pi}{80}[0 + 3(0.3090 + 0.5878 + 0.9511 + 1 + 0.8090 + 0.5878) + 2(0.8090 + 0.9511 + 0.3090)]$$

$$= 0.1178\,[0 + 3(4.2447) + 2(2.0691)]$$

$$\boxed{= 1.9876}$$

Numerical Differentiation

Newton's Forward Difference Formula for Differentiation

(i) First Derivative

$$\left(\frac{dy}{dx}\right)_{x=x_0} = \frac{1}{h}\left[\Delta y_0 - \frac{1}{2}\Delta^2 y_0 + \frac{1}{3}\Delta^3 y_0 - \frac{1}{4}\Delta^4 y_0 + \frac{1}{5}\Delta^5 y_0 - \cdots\right]$$

(ii) Second Derivative

$$\left(\frac{d^2y}{dx^2}\right)_{x=x_0} = \frac{1}{h^2}\left[\Delta^2 y_0 - \Delta^3 y_0 + \frac{11}{12}\Delta^4 y_0 - \frac{5}{6}\Delta^5 y_0 + \cdots\right]$$

Example 1

From the following table of values of $x$and y, obtain

$$\frac{dy}{dx} \text{ and } \frac{d^2y}{dx^2}$$

at $x = 1.2$.

Given Table

x	y
1.0	2.7183
1.2	3.3201
1.4	4.0552
1.6	4.9530
1.8	6.0496
2.0	7.3891
2.2	9.0250

Solution:

Forward Difference Table

x	y	Δ	Δ^2	Δ^3	Δ^4	Δ^5	Δ^6
1.0	2.7183	0.6018	0.1333	0.0294	0.0067	0.0013	0.0001
1.2	3.3201	0.7351	0.1627	0.0361	0.0080	0.0014	
1.4	4.0552	0.8978	0.1988	0.0441	0.0094		
1.6	4.9530	1.0966	0.2429	0.0535			
1.8	6.0496	1.3395	0.2964				
2.0	7.3891	1.6359					
2.2	9.0250						

First Derivative

Using Newton's forward difference formula,

$$h = 1.4 - 1.2 = 0.2$$

$$\left(\frac{dy}{dx}\right)_{x=1.2} = \frac{1}{0.2}\left[0.7351 - \frac{1}{2}(0.1627) + \frac{1}{3}(0.0361) - \frac{1}{4}(0.0080) + \frac{1}{5}(0.0014)\right]$$

$$= 5\,[0.7351 - 0.0814 + 0.0120 - 0.0020 + 0.0003]$$

$$\boxed{\left(\frac{dy}{dx}\right)_{x=1.2} = 3.3205}$$

Second Derivative

$$\left(\frac{d^2y}{dx^2}\right)_{x=1.2} = \frac{1}{0.04}\left[0.1627 - 0.0361 + \frac{11}{12}(0.0080) - \frac{5}{6}(0.0014)\right]$$

$$= 25[0.1627 - 0.0361 + 0.0073 - 0.0011]$$

$$\boxed{\left(\frac{d^2y}{dx^2}\right)_{x=1.2} = 3.32}$$

Example 2

Find $\frac{dy}{dx}, \frac{d^2y}{dx^2}$ at $x = 0$using the following table.

Given

x	0	2	4	6	8	10
$f(x)$	0	12	248	1284	4080	9980

Solution:

Forward Difference Table

x	$y = f(x)$	Δ	Δ^2	Δ^3	Δ^4	Δ^5
0	0	12	224	576	384	0
2	12	236	800	960	384	
4	248	1036	1760	1344		
6	1284	2796	3104			
8	4080	5900				
10	9980					

First Derivative

Using Newton's forward difference formula,

$$h = 2$$

$$\left(\frac{dy}{dx}\right)_{x=0} = \frac{1}{2}\left[12 - \frac{1}{2}(224) + \frac{1}{3}(576) - \frac{1}{4}(384) + 0\right]$$

$$= 0.5[12 - 112 + 192 - 96]$$
$$= 0.5(-4)$$
$$\boxed{\left(\frac{dy}{dx}\right)_{x=0} = -2}$$

Second Derivative

$$\left(\frac{d^2y}{dx^2}\right)_{x=0} = \frac{1}{2^2}\left[224 - 576 + \frac{11}{12}(384)\right]$$
$$= \frac{1}{4}[224 - 576 + 352]$$
$$= 0.25(0)$$
$$\boxed{\left(\frac{d^2y}{dx^2}\right)_{x=0} = 0}$$

Example 3

Find

$$\frac{dy}{dx}, \frac{d^2y}{dx^2}$$

at $x = 1.5$using the following table.

Given

x	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**
$f(x)$	3.375	7.0	13.625	24.0	38.875	59.0

Solution:

Forward Difference Table

x	y	Δ	Δ^2	Δ^3	Δ^4	Δ^5
1.5	3.375	3.625	3.000	0.750	0	0
2.0	7.000	6.625	3.750	0.750	0	
2.5	13.625	10.375	4.500	0.750		
3.0	24.000	14.875	5.250			
3.5	38.875	20.125				
4.0	59.000					

Using Newton's forward difference formula,

$$h = 2.0 - 1.5 = 0.5$$

$$\left(\frac{dy}{dx}\right)_{x=1.5} = \frac{1}{0.5}\left[3.625 - \frac{1}{2}(3) + \frac{1}{3}(0.75)\right]$$

(The remaining part of the solution is cut off in the provided image.)

First Derivative

$$\left(\frac{dy}{dx}\right)_{x=x_0} = \frac{1}{h}\left[\Delta y_0 - \frac{1}{2}\Delta^2 y_0 + \frac{1}{3}\Delta^3 y_0 - \frac{1}{4}\Delta^4 y_0 + \cdots\right]$$

$$= \frac{1}{0.5}\left[3.625 - \frac{1}{2}(3) + \frac{1}{3}(0.75)\right]$$

$$= 2(3.625 - 1.5 + 0.25)$$

$$= 2(2.375)$$

$$= 4.75$$

Second Derivative

$$\left(\frac{d^2y}{dx^2}\right)_{x=x_0} = \frac{1}{h^2}\left[\Delta^2 y_0 - \Delta^3 y_0 + \frac{11}{12}\Delta^4 y_0 - \frac{5}{6}\Delta^5 y_0 + \cdots\right]$$

$$= \frac{1}{(0.5)^2}(3 - 0.75)$$

$$= \frac{2.25}{0.25} = 9$$

$$\boxed{\left(\frac{d^2y}{dx^2}\right)_{x=1.5} = 9}$$

Newton's Backward Formula for Differentiation

(i) First Derivative

$$\left(\frac{dy}{dx}\right)_{x=x_n} = \frac{1}{h}\left[\nabla y_n + \frac{1}{2}\nabla^2 y_n + \frac{1}{3}\nabla^3 y_n + \frac{1}{4}\nabla^4 y_n + \cdots\right]$$

(ii) Second Derivative

$$\left(\frac{d^2y}{dx^2}\right)_{x=x_n} = \frac{1}{h^2}\left[\nabla^2 y_n + \nabla^3 y_n + \frac{11}{12}\nabla^4 y_n + \frac{5}{6}\nabla^5 y_n + \cdots\right]$$

Example 4

Find

$$\frac{dy}{dx}, \frac{d^2y}{dx^2}$$

for $x = 2.2$.

Given Table

x	**1.0**	**1.2**	**1.4**	**1.6**	**1.8**	**2.0**	**2.2**
y	2.7183	3.3201	4.0552	4.9530	6.0496	7.3891	9.0250

Solution:

Backward Difference Table:

x	y	∇	∇^2	∇^3	∇^4	∇^5	∇^6
1.0	2.7183	0.6018	0.1333	0.0294	0.0067	0.0013	0.0001
1.2	3.3201	0.7351	0.1627	0.0361	0.0080	0.0014	
1.4	4.0552	0.8978	0.1988	0.0441	0.0094		
1.6	4.9530	1.0966	0.2429	0.0535			
1.8	6.0496	1.3395	0.2964				
2.0	7.3891	1.6359					
2.2	9.0250						

Given,

$$h = 0.2$$

(i) First Derivative

$$\left(\frac{dy}{dx}\right)_{x=2.2} = \frac{1}{0.2}\left[1.6359 + \frac{1}{2}(0.2964) + \frac{1}{3}(0.0535) + \frac{1}{4}(0.0094) + \frac{1}{5}(0.0014) + \frac{1}{6}(0.0001)\right]$$

$$= 5(1.6359 + 0.1482 + 0.0178 + 0.00235 + 0.00028 + 0.00001)$$

$$\boxed{\left(\frac{dy}{dx}\right)_{x=2.2} = 9.022}$$

(ii) Second Derivative

$$\left(\frac{d^2y}{dx^2}\right)_{x=2.2} = \frac{1}{0.04}\left[0.2964 + 0.0535 + \frac{11}{12}(0.0094) + \frac{5}{6}(0.0014) + \frac{137}{180}(0.0001)\right]$$

$$= 25(0.2964 + 0.0535 + 0.0086 + 0.0011 + 0)$$

$$\boxed{\left(\frac{d^2y}{dx^2}\right)_{x=2.2} = 8.99}$$

Example 5

Find

$$\frac{dy}{dx}, \frac{d^2y}{dx^2}$$

when $x = 6$.

Given Table

x	**0**	**1**	**2**	**3**	**4**	**5**	**6**
y	6.9897	7.4036	7.7815	8.1291	8.4510	8.7506	9.0309

Solution:

Backward Difference Table:

x	y	∇	∇^2	∇^3	∇^4	∇^5	∇^6
0	6.9897	0.4139	-0.0360	0.0057	-0.0014	0.0005	0
1	7.4036	0.3779	-0.0303	0.0043	-0.0009	0.0005	
2	7.7815	0.3476	-0.0257	0.0034	-0.0004		
3	8.1291	0.3219	-0.0223	0.0030			
4	8.4510	0.2996	-0.0193				
5	8.7506	0.2803					
6	9.0309						

Given,

$$h = 6 - 5 = 1$$

(i) First Derivative

$$\left(\frac{dy}{dx}\right)_{x=6} = \frac{1}{h}\left[\nabla y_n + \frac{1}{2}\nabla^2 y_n + \frac{1}{3}\nabla^3 y_n + \frac{1}{4}\nabla^4 y_n + \frac{1}{5}\nabla^5 y_n\right]$$

$$= \frac{1}{1}\left[0.2803 + \frac{1}{2}(-0.0193) + \frac{1}{3}(0.0030) + \frac{1}{4}(-0.0004) + \frac{1}{5}(0.0005)\right]$$

$$= 0.2803 - 0.00965 + 0.0010 - 0.0001 + 0.0001$$

$$\boxed{\left(\frac{dy}{dx}\right)_{x=6} = 0.27165}$$

Numerical Differentiation

6. Estimate the Rate of Growth of the Population in the Year 1981 from the Following Table

Given:

Year (x)	**1951**	**1961**	**1971**	**1981**	**1991**
Population (y)	19.96	39.65	58.81	77.21	94.61

SOLUTION:

Difference Table

X	**Y**	**∇**	**∇²**	**∇³**	**∇⁴**
1951	19.96				
1961	39.65	19.69	-0.53	-0.23	-0.01
1971	58.81	19.16	-0.76	-0.24	
1981	77.21	18.40	-1.00		
1991	94.61	17.40			

Here,

$$h = 1991 - 1981 = 10$$

Using Newton's Backward Formula,

$$\left(\frac{dy}{dx}\right)_{x=x_n} = \frac{1}{h}\left[\nabla y_n + \frac{1}{2}\nabla^2 y_n + \frac{1}{3}\nabla^3 y_n + \frac{1}{4}\nabla^4 y_n + \cdots\right]$$

$$= \frac{1}{10}\left[18.4 + \frac{1}{2}(-0.76) + \frac{1}{3}(-0.23)\right]$$

$$= 0.1(18.4 - 0.38 - 0.076)$$

$$= 0.1(17.9434)$$

$$= 1.79434$$

Therefore,

$$\boxed{\left(\frac{dy}{dx}\right)_{x=198} = 1.79434}$$

Using Newton's Backward Formula for Second Derivative,

$$\left(\frac{d^2y}{dx^2}\right)_{x=x_n} = \frac{1}{h^2}\left[\nabla^2 y_n + \nabla^3 y_n + \frac{11}{12}\nabla^4 y_n + \frac{5}{6}\nabla^5 y_n + \cdots\right]$$

Here,

$$h = 10, h^2 = 100$$
$$= \frac{1}{100}\left[-1 + (-0.24) + \frac{11}{12}(-0.23)\right]$$
$$= -0.01429$$

Hence,

$$\boxed{\left(\frac{d^2y}{dx^2}\right)_{x=198} = -0.01429}$$

Least Squares Curve Fitting Procedure

The method of **Least Squares** is probably the most systematic procedure to fit a unique curve through the given data points and is widely used in practical computations. It can also be easily implemented on a digital computer.

Let the set of data points be

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

and let the curve

$$y = f(x)$$

be fitted to these data.

If $e_i$is the error of approximation at $x = x_i$, then

$$e_1 = y_1 - f(x_1)$$
$$e_2 = y_2 - f(x_2)$$
$$\vdots$$
$$e_n = y_n - f(x_n)$$

Hence,

$$S = [y_1 - f(x_1)]^2 + [y_2 - f(x_2)]^2 + \cdots + [y_n - f(x_n)]^2$$

or

$$S = e_1^2 + e_2^2 + \cdots + e_n^2$$

The method of least squares consists of minimizing the quantity S.

Derivation of Normal Equations for Least Square Approximation

Let the straight line to be fitted be

$$y = a_0 + a_1 x$$

Then,

$$S = [y_1 - (a_0 + a_1 x_1)]^2 + [y_2 - (a_0 + a_1 x_2)]^2 + \cdots + [y_m - (a_0 + a_1 x_m)]^2$$

For minimum error,

$$\frac{dS}{da_0} = 0$$

and

$$\frac{dS}{da_1} = 0$$

Differentiating,

$$\frac{dS}{da_0} = -2[y_1 - (a_0 + a_1 x_1)] - 2[y_2 - (a_0 + a_1 x_2)] - \cdots - 2[y_m - (a_0 + a_1 x_m)] = 0$$

Similarly,

$$\frac{dS}{da_1} = -2x_1[y_1 - (a_0 + a_1 x_1)] - 2x_2[y_2 - (a_0 + a_1 x_2)] - \cdots - 2x_m[y_m - (a_0 + a_1 x_m)] = 0$$

These simplify to

First Normal Equation

$$m a_0 + a_1 \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i$$

Second Normal Equation

$$a_0 \sum_{i=1}^{m} x_i + a_1 \sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} x_i y_i$$

Equations (1) and (2) are called the **Normal Equations** for the least squares straight-line fit.

Least Squares Method

Example 1

Fit a Straight Line to the Following Data and Find $y$when $x = 6$

Given Data

x	0	5	10	15	20	25
y	12	15	17	22	24	30

Solution

Let the straight line be

$$y = a_0 + a_1 x$$

Construct the table:

x_i	y_i	x_i^2	$x_i y_i$
0	12	0	0
5	15	25	75
10	17	100	170
15	22	225	330
20	24	400	480
25	30	625	750

$$\sum x_i = 75, \sum y_i = 120$$
$$\sum x_i^2 = 1375, \sum x_i y_i = 1805$$

Here,

$$m = 6$$

First Normal Equation

$$m a_0 + a_1 \sum x_i = \sum y_i$$
$$6a_0 + 75a_1 = 120 \qquad (1)$$

Second Normal Equation

$$a_0 \sum x_i + a_1 \sum x_i^2 = \sum x_i y_i$$
$$75a_0 + 1375a_1 = 1805 \qquad (2)$$

Multiply (1) by 75:

$$450a_0 + 5625a_1 = 9000$$

Multiply (2) by 6:

$$450a_0 + 8250a_1 = 10830$$

Subtracting,

$$2625a_1 = 1830$$
$$a_1 = \frac{1830}{2625} = 0.697$$

Substitute into (1):

$$6a_0 + 75(0.697) = 120$$
$$6a_0 = 67.725$$
$$a_0 = \frac{67.725}{6} = 11.28$$

Hence the required straight line is

$$\boxed{y = 11.28 + 0.697x}$$

For $x = 6$,

$$y = 11.28 + 0.697(6)$$
$$= 11.28 + 4.182$$
$$\boxed{y = 15.462}$$

Example 2

The Following Table Gives the Temperature T(°C) and the Length l(mm) of a Heated Rod.
If

$$l = a_0 + a_1T,$$

find the best values of $a_0$and a_1.

Given Data

T(°C)	20	30	40	50	60	70
l(mm)	800.3	800.4	800.6	800.7	800.9	801.0

Solution

Let

$$l = a_0 + a_1T$$

Prepare the table.

T_i	l_i	T_i^2	T_il_i
20	800.3	400	16006
30	800.4	900	24012
40	800.6	1600	32024
50	800.7	2500	40035
60	800.9	3600	48054
70	801.0	4900	56070

$$\sum T_i = 270$$
$$\sum l_i = 4803.9$$
$$\sum T_i^2 = 13900$$
$$\sum T_il_i = 216201$$

First Normal Equation

$$6a_0 + 270a_1 = 4803.9 \qquad (1)$$

Second Normal Equation

$$270a_0 + 13900a_1 = 216201 \qquad (2)$$

Multiply (1) by 270:

$$1620a_0 + 72900a_1 = 1297053$$

Multiply (2) by 6:

$$1620a_0 + 83400a_1 = 1297206$$

Subtracting,

$$10500a_1 = 153$$
$$a_1 = \frac{153}{10500} = 0.01457 \approx 0.0146$$

Substitute into (1):

$$6a_0 + 270(0.0146) = 4803.9$$
$$6a_0 + 3.942 = 4803.9$$
$$6a_0 = 4799.958$$
$$a_0 \approx 799.99 \approx 800$$

Hence,

$$\boxed{a_0 \approx 800, a_1 \approx 0.0146}$$

Example 3

Certain Experimental Values of $x$and $y$are Given Below.

If

$$y = a_0 + a_1 x,$$

find the approximate values of $a_0$and a_1.

Given Data

x	**0**	**2**	**5**	**7**
y	-1	5	12	20

Solution

Let

$$y = a_0 + a_1 x$$

Construct the table.

x_i	y_i	x_i^2	x_iy_i
0	-1	0	0
2	5	4	10
5	12	25	60
7	20	49	140

$$\sum x_i = 14$$
$$\sum y_i = 36$$
$$\sum x_i^2 = 78$$
$$\sum x_iy_i = 210$$

Here,

$$m = 4$$

First Normal Equation

$$4a_0 + 14a_1 = 36 \qquad (1)$$

Second Normal Equation

$$14a_0 + 78a_1 = 210 \qquad (2)$$

Multiply (1) by 14:

$$56a_0 + 196a_1 = 504$$

Multiply (2) by 4:

$$56a_0 + 312a_1 = 840$$

Subtracting,

$$116a_1 = 336$$
$$a_1 = \frac{336}{116} = 2.89$$

Substitute into (1):

$$4a_0 + 14(2.89) = 36$$
$$4a_0 = 36 - 40.46 = -4.46$$
$$a_0 = -1.115$$

Hence,

$$\boxed{a_0 = -1.115, a_1 = 2.89}$$

Example 4

By the Method of Least Squares, Find the Straight Line that Best Fits the Following Data.

Given Data

x	1	2	3	4	5
y	14	27	40	55	68

Solution

Let the straight line be

$$y = a_0 + a_1 x$$

Construct the table.

x_i	y_i	x_i^2	$x_i y_i$
1	14	1	14
2	27	4	54
3	40	9	120
4	55	16	220
5	68	25	340

$$\sum x_i = 15$$
$$\sum y_i = 204$$
$$\sum x_i^2 = 55$$
$$\sum x_i y_i = 748$$

From (1) and (2):

$$75 \times (1) \Rightarrow 4500a_0 + 5625a_1 = 9000$$
$$6 \times (2) \Rightarrow 4500a_0 + 8250a_1 = 10830$$

Subtracting,

$$2625a_1 = 1830$$
$$a_1 = \frac{1830}{2625} = 0.697$$

Substituting in (1),

$$6a_0 + 75(0.697) = 120$$
$$6a_0 = 120 - 52.275 = 67.725$$
$$a_0 = \frac{67.725}{6} = 11.28$$

Hence, the straight line passing through the given points is

$$\boxed{y = 11.28 + 0.697x}$$

For $x = 6$,

$$y = 11.28 + 0.697(6) = 15.462 \approx 15.468$$

---

2. The table below gives the temperature T(in °C) and length l(in mm) of a heated rod. If

$$l = a_0 + a_1 T,$$

find the best values of $a_0$and a_1.

Given Data

T(°C)	l(mm)
20	800.3
30	800.4
40	800.6
50	800.7
60	800.9
70	801.0

Solution

Let the straight line be

$$l = a_0 + a_1 T$$

T_i	l_i	T_i^2	$T_i l_i$
20	800.3	400	16006
30	800.4	900	24012
40	800.6	1600	32024
50	800.7	2500	40035
60	800.9	3600	48054
70	801.0	4900	56070

$$\sum T_i = 270, \sum l_i = 4803.9,$$
$$\sum T_i^2 = 13900, \sum T_i l_i = 216201$$

First Normal Equation

$$6a_0 + 270a_1 = 4803.9 \qquad (1)$$

Second Normal Equation

$$270a_0 + 13900a_1 = 216201 \qquad (2)$$

From (1) and (2),

$$(1) \times 270:$$
$$1620a_0 + 72900a_1 = 1297053$$
$$(2) \times 6:$$
$$1620a_0 + 83400a_1 = 1297206$$

Subtracting,

$$10500a_1 = 153$$
$$a_1 = \frac{153}{10500} = 0.01457 \approx 0.0146$$

Substituting in (1),

$$6a_0 + 270(0.0146) = 4803.9$$
$$6a_0 = 4803.9 - 3.942 = 4799.958$$
$$a_0 = \frac{4799.958}{6} \approx 799.99 \approx 800$$

Hence,

$$\boxed{a_0 \approx 800, a_1 \approx 0.0146}$$

Y=800+0.0146X

---

3. Certain experimental values of $x$and $y$are given below. If

$$y = a_0 + a_1x,$$

find the approximate values of $a_0$and a_1.

Given Data

x	**0**	**2**	**5**	**7**
y	-1	5	12	20

Solution

Let the straight line be

$$y = a_0 + a_1x$$

x_i	y_i	x_i^2	x_iy_i
0	-1	0	0
2	5	4	10
5	12	25	60
7	20	49	140

$$\sum x_i = 14, \sum y_i = 36,$$

$$\sum x_i^2 = 78, \sum x_i y_i = 210$$

Here,

$$m = 4$$

First Normal Equation

$$4a_0 + 14a_1 = 36 \qquad (1)$$

Second Normal Equation

$$14a_0 + 78a_1 = 210 \qquad (2)$$

From (1) and (2),

$$(1) \times 14:$$
$$56a_0 + 196a_1 = 504$$
$$(2) \times 4:$$
$$56a_0 + 312a_1 = 840$$

Subtracting,

$$116a_1 = 336$$
$$a_1 = \frac{336}{116} \approx 2.89$$

Substituting in (1),

$$4a_0 + 14(2.89) = 36$$
$$4a_0 = 36 - 40.46 = -4.46$$
$$a_0 = -1.115$$

Hence,

$$\boxed{a_0 \approx -1.115, a_1 \approx 2.89}$$
$$Y = -1.115 + 2.89X$$

---

4. By the method of Least Squares, find the straight line that best fits the following data.

Given Data

x	**1**	**2**	**3**	**4**	**5**
y	14	27	40	55	68

Solution

Let the straight line be

$$y = a_0 + a_1 x$$

x_i	y_i	x_i^2	$x_i y_i$
1	14	1	14
2	27	4	54
3	40	9	120
4	55	16	220
5	68	25	340

$$\sum x_i = 15, \sum y_i = 204,$$
$$\sum x_i^2 = 55, \sum x_i y_i = 748$$

Here,

$$m = 5$$

First Normal Equation

$$5a_0 + 15a_1 = 204 \qquad (1)$$

Second Normal Equation

$$15a_0 + 55a_1 = 748 \qquad (2)$$

From (1) and (2),

$$(1) \times 15:$$
$$75a_0 + 225a_1 = 3060$$
$$(2) \times 5:$$
$$75a_0 + 275a_1 = 3740$$

Subtracting,

$$50a_1 = 680$$
$$a_1 = 13.6$$

Substituting in (1),

$$5a_0 + 15(13.6) = 204$$
$$5a_0 = 0$$
$$a_0 = 0$$

Hence, the required straight line is

$$\boxed{y = 13.6x}$$

Fit a Polynomial of Second Degree:

Let

$$\boxed{y = a_0 + a_1 x + a_2 x^2}$$

(parabola)

Normal Equations

1. First Normal Equation

$$ma_0 + \sum x_i a_1 + \sum x_i^2 a_2 = \sum y_i$$

2. Second Normal Equation

$$\sum x_i a_0 + \sum x_i^2 a_1 + \sum x_i^3 a_2 = \sum x_i y_i$$

3. Third Normal Equation

$$\sum x_i^2 a_0 + \sum x_i^3 a_1 + \sum x_i^4 a_2 = \sum x_i^2 y_i$$

Problem 1

Find the best values of a_0, a_1, a_2 so that the parabola

$$y = a_0 + a_1 x + a_2 x^2$$

fits the data:

x	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**
y	1.1	1.2	1.5	2.6	2.8	3.3	4.1

SOLUTION

Computation Table

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
1.0	1.1	1.000	1.000	1.000	1.10	1.10
1.5	1.2	2.250	3.375	5.0625	1.80	2.70
2.0	1.5	4.000	8.000	16.000	3.00	6.00
2.5	2.6	6.250	15.625	39.0625	6.50	16.25
3.0	2.8	9.000	27.000	81.000	8.40	25.20
3.5	3.3	12.250	42.875	150.0625	11.55	40.425
4.0	4.1	16.000	64.000	256.000	16.40	65.60
Σ	**16.6**	**50.75**	**161.875**	**548.125**	**48.75**	**157.275**

Normal Equations

Using the above totals,

First Normal Equation

$$7a_0 + 17.5a_1 + 50.75a_2 = 16.6$$

Second Normal Equation

$$17.5a_0 + 50.75a_1 + 161.875a_2 = 48.75$$

Third Normal Equation

$$50.75a_0 + 161.875a_1 + 548.125a_2 = 157.275$$

Solving the Equations

After elimination,

$$49a_1 + 245a_2 = 50.750$$
$$245a_1 + 1261.75a_2 = 285.44$$

Multiplying the first equation by 5,

$$245a_1 + 1225a_2 = 253.75$$

Subtracting,

$$36.75a_2 = 4.69$$
$$\boxed{a_2 = 0.1276}$$

Now,

$$49a_1 + 245(0.1276) = 50.750$$
$$49a_1 = 19.488$$
$$\boxed{a_1 = 0.3929}$$

Using the first normal equation,

$$7a_0 = 16.6 - 17.5(0.3929) - 50.75(0.1276)$$
$$7a_0 = 3.1976$$
$$\boxed{a_0 = 0.4568}$$

Best-Fit Polynomial

$$\boxed{y = 0.4568 + 0.3929x + 0.1276x^2}$$

This is the required second-degree polynomial (parabola) that best fits the given data.

2. Fit a Polynomial of the Second Degree

Fit a polynomial of the second degree to the data given in the following table.

x	**0.0**	**1.0**	**2.0**
y	1.0	6.0	17.0

Assume

$$y = a_0 + a_1x + a_2x^2$$

Computation Table

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
0	1	0	0	0	0	0
1	6	1	1	1	6	6
2	17	4	8	16	34	68
Σ	**24**	**5**	**9**	**17**	**40**	**74**

Normal Equations

First Normal Equation

$$3a_0 + 3a_1 + 5a_2 = 24(1)$$

Second Normal Equation

$$3a_0 + 5a_1 + 9a_2 = 40(2)$$

Third Normal Equation

$$5a_0 + 9a_1 + 17a_2 = 74(3)$$

Solution

From (2) – (1),

$$2a_1 + 4a_2 = 16$$
$$a_1 + 2a_2 = 8(4)$$

Multiply (2) by 5:

$$15a_0 + 25a_1 + 45a_2 = 200$$

Multiply (3) by 3:

$$15a_0 + 27a_1 + 51a_2 = 222$$

Subtracting,

$$2a_1 + 6a_2 = 22$$
$$a_1 + 3a_2 = 11(5)$$

Subtract (4) from (5):

$$a_2 = 3$$

Substitute into (4):

$$a_1 + 2(3) = 8$$
$$a_1 = 2$$

Substitute $a_1 = 2$and $a_2 = 3$into (1):

$$3a_0 + 3(2) + 5(3) = 24$$
$$3a_0 = 3$$
$$a_0 = 1$$

Required Polynomial

$$\boxed{y = a_0 + a_1x + a_2x^2}$$
$$\boxed{y = 1 + 2x + 3x^2}$$

Problem 3

Find the best values of a, b, and $c$so that the parabola

$$y = a + bx + cx^2$$

fits the following data.

x	**0**	**1**	**2**	**3**	**4**
y	1	0	3	10	21

Computation Table

x_i	y_i	x_i^2	x_i^3	x_i^4	x_iy_i	$x_i^2y_i$
0	1	0	0	0	0	0
1	0	1	1	1	0	0
2	3	4	8	16	6	12
3	10	9	27	81	30	90
4	21	16	64	256	84	336
Σ	**35**	**30**	**100**	**354**	**120**	**438**

Hence, the required polynomial is assumed as

$$\boxed{y = a + bx + cx^2}$$

Normal Equations

First Normal Equation

$$5a + 10b + 30c = 35(1)$$

Second Normal Equation

$$10a + 30b + 100c = 120(2)$$

Third Normal Equation

$$30a + 100b + 354c = 438(3)$$

Solution

From equations (1) and (2):

Multiply (1) by 10:

$$50a + 100b + 300c = 350$$

Multiply (2) by 5:

$$50a + 150b + 500c = 600$$

Subtracting,

$$-50b - 200c = -250$$
$$b + 4c = 5(4)$$

From equations (2) and (3):

Multiply (2) by 10:

$$100a + 300b + 1000c = 1200$$

Multiply (3) by 3:

$$90a + 300b + 1062c = 1314$$

Subtracting,

$$10a - 62c = -114$$

Using the elimination shown in the notes,

$$100b + 540c = 780$$
$$5b + 27c = 39(5)$$

From equations (4) and (5):

Multiply (4) by 5:

$$5b + 20c = 25$$

Subtract (5):

$$(5b + 27c) - (5b + 20c) = 39 - 25$$

$$7c = 14$$
$$\boxed{c = 2}$$

Substitute into equation (4):

$$b + 4(2) = 5$$
$$b = -3$$

Substitute $b = -3$and $c = 2$ into equation (1):

$$5a + 10(-3) + 30(2) = 35$$
$$5a - 30 + 60 = 35$$
$$5a = 5$$
$$\boxed{a = 1}$$

Required Polynomial

$$\boxed{y = a + bx + cx^2}$$
$$\boxed{y = 1 - 3x + 2x^2}$$

PRACTICE QUESTIONS

Question 1

Fit a second-degree polynomial to the following data:

x	0	1	2	3	4
y	1	2	5	10	17

Question 2

Fit a parabola

y=a+bx+cx
2

to the following observations.

x	1	2	3	4	5
y	2	5	10	17	26

Question 3

Fit a second-degree polynomial for the data.

x	0	2	4	6	8
y	3	7	15	27	43

Question 4

Fit a parabola using the least squares method.

x	1	2	3	4
y	4	7	12	19

PROBLEM 5.

Using the second-order Runge–Kutta method, find $y(0.2)$if

$$\frac{dy}{dx} = x + y, y(0) = 1,$$

taking $h = 0.2$.

PROBLEM 6

Using the Trapezoidal Rule, evaluate

$$\int_1^3 (x^3 + 2x)\, dx$$

taking $n = 4$.

PROBLEM 7. Using Simpson's $\frac{1}{3}$Rule, evaluate

$$\int_0^2 (x^3 + 1)\, dx$$

taking four equal intervals.

PROBLEM 8. Using Simpson's $\frac{3}{8}$Rule, evaluate

$$\int_0^3 (x^2 + 2x)\, dx.$$

PROBLEM 9. Using Newton's Forward Interpolation Formula, find $y(23)$.

x	**20**	**22**	**24**	**26**	**28**
y	142	183	228	277	330

PROBLEM 10. Using Newton's Forward Interpolation, find $y(18)$.

x	**15**	**17**	**19**	**21**	**23**
y	225	289	361	441	529

PROBLEM 11. Using Newton's Backward Interpolation Formula, find $y(37)$.

x	**25**	**30**	**35**	**40**	**45**
y	112	148	190	238	292

UNIT–3
Solution of Initial Value Problem to Ordinary Differential Equations

UNIT–3
Solution of Initial Value Problem to Ordinary Differential Equations

Numerical Solution of Ordinary Differential Equations

Many problems in Science and Engineering can be formulated into ordinary differential equations. The analytical methods of solving differential equations are applicable only to a selected class of differential equations. Quite often, equations occurring in physical problems do not belong to any of these simple types. Hence, numerical methods are used for solving such differential equations.

The following numerical methods are discussed:

1. Taylor's Series Method
2. Picard's Method
3. Euler's Method
4. Runge–Kutta Method

---

1. Taylor's Series Method

To find the numerical solution of the differential equation

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

with the initial condition

$$y(x_0) = y_0$$

Method

Differentiate the given equation repeatedly to obtain

$$y', y'', y''', y^{(4)}, \ldots$$

Substitute the values at

$$x = x_0, y = y_0$$

into Taylor's expansion:

$$\boxed{y(x) = y_0 + (x - x_0)y_0' + \frac{(x - x_0)^2}{2!}y_0'' + \frac{(x - x_0)^3}{3!}y_0''' + \cdots}$$

Thus, Taylor's series gives the approximate value of $y$for nearby values of x.

Example 1

Solve

$$\frac{dy}{dx} = xy + 1, y(0) = 1$$

using Taylor's method and compute $y(0.1)$.

Solution

Given

$$y' = xy + 1$$

Differentiate successively:

$$\begin{aligned} y'' &= y + xy' \\ y''' &= 2y' + xy'' \\ y^{(4)} &= 3y'' + xy''' \end{aligned}$$

At

$$x = 0, y = 1$$

we obtain

$$\begin{aligned} y_0' &= 1 \\ y_0'' &= 1 \\ y_0''' &= 2 \\ y_0^{(4)} &= 3 \end{aligned}$$

Hence,

$$y(x) = 1 + x + \frac{x^2}{2!} + \frac{2x^3}{3!} + \frac{3x^4}{4!} + \cdots$$

At $x = 0.1$,

$$\begin{aligned} y(0.1) &= 1 + 0.1 + \frac{0.01}{2} + \frac{2(0.001)}{6} + \frac{3(0.0001)}{24} \\ &= 1 + 0.1 + 0.005 + 0.000333 + 0.0000125 \\ &\approx \boxed{1.1053} \end{aligned}$$

---

2. Picard's Method of Successive Approximation

Consider the initial value problem

$$\frac{dy}{dx} = f(x, y), y(x_0) = y_0$$

Method

Integrate both sides:

$$dy = f(x, y)\, dx$$

Applying the limits,

$$\boxed{y(x) = y_0 + \int_{x_0}^{x} f(x, y)\, dx}$$

Since $y$appears inside the integral, successive approximations are used.
The Picard iteration formula is

$$\boxed{y_{n+1} = y_0 + \int_{x_0}^{x} f(t, y_n)\, dt}$$

where

First approximation

$$y_1 = y_0 + \int_{x_0}^{x} f(t, y_0)\, dt$$

Second approximation

$$y_2 = y_0 + \int_{x_0}^{x} f(t, y_1)\, dt$$

Similarly,

$$y_3,\ y_4, \dots$$

are obtained until the desired accuracy is reached.

Example 1

Using Picard's method, find the solution of

$$\frac{dy}{dx} = 1 + xy, y(0) = 0$$

up to the third approximation.

Solution

Picard iteration formula:

$$y_{n+1} = 0 + \int_{0}^{x} (1 + t\ y_n)\, dt$$

First approximation

Take

$$y_0 = 0$$

Then

$$y_1 = \int_0^x 1\,dt = x$$

Second approximation

$$y_2 = \int_0^x (1 + t\,y_1)\,dt = \int_0^x (1 + t^2)\,dt = x + \frac{x^3}{3}$$

Third approximation

$$y_3 = \int_0^x \left(1 + t\left(t + \frac{t^3}{3}\right)\right) dt = \int_0^x \left(1 + t^2 + \frac{t^4}{3}\right) dt = x + \frac{x^3}{3} + \frac{x^5}{15}$$

Hence,

$$\boxed{y_3 = x + \frac{x^3}{3} + \frac{x^5}{15}}$$

EXAMPLE 2

Using Picard's method, find the value of y at $x = 0.4$ for

$$\frac{dy}{dx} = x^2 + y^2, y(0) = 0.$$

Solution:

Given

$$\frac{dy}{dx} = x^2 + y^2, y(0) = 0, x_0 = 0$$
$$[\therefore \ y(x_0) = y_0]$$

We know that Picard's iteration formula is

$$y_n = y_0 + \int_{x_0}^{x} f(x, y_{n-1})\, dx (i)$$

Let

1st approximation

Put $n = 1$in Eq. (i), we get

$$y_1 = y_0 + \int_{x_0}^{x} f(x, y_0)\, dx$$

Here,

$$f(x, y) = x^2 + y^2$$
$$f(x, y_0) = x^2$$

Hence,

$$y_1 = 0 + \int_0^x x^2 \ dx$$
$$= \left[\frac{x^3}{3}\right]_0^x$$
$$= \frac{x^3}{3}$$

Also,

$$f(x, y_1) = x^2 + \left(\frac{x^3}{3}\right)^2 = x^2 + \frac{x^6}{9}$$

2nd approximation

$$y_2 = y_0 + \int_{x_0}^{x} f(x, y_1)\, dx$$
$$= 0 + \int_0^x \left(x^2 + \frac{x^6}{9}\right) dx$$
$$= \frac{x^3}{3} + \frac{x^7}{63}$$

Calculation of y_3 is difficult and hence the approximate value is taken as y_2.
For $x = 0.4$,

$$y_2 = \frac{(0.4)^3}{3} + \frac{(0.4)^7}{63}$$
$$= 0.02136 + 0.00003$$
$$\boxed{\approx 0.02139}$$

Example 3

Find the value of y for $x = 0.25,\ 0.5,\ 1$ by Picard's method

Given that

$$\frac{dy}{dx} = \frac{x^2}{y^2+1}, y(0) = 0$$

Solution:

Given

$$\frac{dy}{dx} = \frac{x^2}{y^2+1}, y(0) = 0, x_0 = 0$$

We know Picard's iteration formula

$$y_n = y_0 + \int_{x_0}^{x} f(x, y_{n-1})\, dx$$

1st approximation

$$y_1 = y_0 + \int_0^x f(x, y_0)\, dx$$

Since

$$f(x, y_0) = \frac{x^2}{0^2+1} = x^2$$

Therefore,

$$y_1 = 0 + \int_0^x x^2\ dx$$
$$= \left[\frac{x^3}{3}\right]_0^x$$
$$= \frac{x^3}{3}$$

Now,

$$f(x, y_1) = \frac{x^2}{\left(\frac{x^3}{3}\right)^2 + 1} = \frac{x^2}{\frac{x^6}{9} + 1} = \frac{9x^2}{x^6 + 9}$$

2nd approximation

$$y_2 = y_0 + \int_0^x f(x, y_1)\, dx$$

$$= 0 + \int_0^x \frac{9x^2}{x^6 + 9}\, dx$$

Let

$$x^3 = t$$

Then

$$3x^2\, dx = dt$$

When $x = 0,\ t = 0$
When $x = x,\ t = x^3$
Hence,

$$= 3\int_0^{x^3} \frac{dt}{t^2 + 3^2}$$

Using

$$\int \frac{dx}{x^2 + a^2} = \frac{1}{a}\tan^{-1}\left(\frac{x}{a}\right)$$

we get

$$= 3\left[\frac{1}{3}\tan^{-1}\left(\frac{t}{3}\right)\right]_0^{x^3}$$

$$= \tan^{-1}\left(\frac{x^3}{3}\right)$$

Therefore,

$$\boxed{y_2 = \tan^{-1}\left(\frac{x^3}{3}\right)}$$

Hence, we use $y_2$as the value of y.

Values

$$y(0.25) = \tan^{-1}\left(\frac{(0.25)^3}{3}\right) \approx 0.0052$$
$$y(0.5) = \tan^{-1}\left(\frac{(0.5)^3}{3}\right) \approx 0.0416$$
$$y(1) = \tan^{-1}\left(\frac{1}{3}\right) \approx 0.3218$$

Euler's Method

We have so far discussed the methods which yield the solution of the differential equation in the form of a function.

We will now describe the method which gives the solution in the form of a set of tabulated values. Suppose we wish to solve the equation

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

given

$$y(x_0) = y_0.$$

In this method, starting from x_0, we take equally spaced values of x, say

$$x_0,\ x_1,\ x_2,\ \ldots,\ x_n$$

with step length h.

Therefore,

$$x_n = x_{n-1} + h, n = 1,2, \ldots$$

The successive approximation values of $y$are given by

$$y(x_1) = y_1 = y_0 + h\,f(x_0, y_0)$$
$$y(x_2) = y_2 = y_1 + h\,f(x_1, y_1)$$
$$y(x_3) = y_3 = y_2 + h\,f(x_2, y_2)$$
$$\vdots$$
$$\boxed{y_{n+1} = y_n + h\,f(x_n, y_n), n = 0,1,2, \ldots}$$

This is known as **Euler's algorithm** and can be used to evaluate

$$y_1,\ y_2,\ \ldots,\ y_n,$$

i.e.,

$$y(x_1),\ y(x_2),\ \ldots,\ y(x_n),$$

starting from the initial condition.

Example 1
Using Euler's method solve for $y$at $x = 2$from

$$\frac{dy}{dx} = 3x^2 + 1, y(1) = 2,$$

taking step size

1. $h = 0.5$
2. $h = 0.25$

Solution
Given

$$\frac{dy}{dx} = 3x^2 + 1$$
$$f(x, y) = 3x^2 + 1, y(1) = 2$$

Here,

$$x_0 = 1, y_0 = 2.$$

We know Euler's algorithm is

$$y_{n+1} = y_n + h\,f(x_n, y_n) \dots\dots\dots\dots \quad (1)$$

Clearly,

$$x_n = x_0 + nh, x_0 = 1.$$

Taking $n = 0$in Eq. (1), we have

$$y_1 = y_0 + h\,f(x_0, y_0)$$
$$= 2 + 0.5\,[\,3(1)^2 + 1\,]$$
$$= 2 + 0.5(4)$$
$$= 4.$$

Also,

$$f(1,2) = 3(1)^2 + 1 = 4.$$

Taking $n = 1$in Eq. (1), we have

$$y_2 = y_1 + h\,f(x_1, y_1)$$
$$= 4 + 0.5\,[\,3(1.5)^2 + 1\,] = 7.8750$$

Taking $n = 2$in Eq. (1), we have

$$
\begin{aligned}
y_3 &= y_2 + h\,f(x_2, y_2) \\
&= 7.8750 + 0.5(3(2)^2 + 1) \\
&= 7.8750 + 0.5(13) \\
&= 14.375
\end{aligned}
$$

Given $h = 0.25$

$$x_0 = 1, x_1 = 1.25, x_2 = 1.5, x_3 = 1.75$$

Taking $n = 0$in Eq. (1), we have

$$
\begin{aligned}
y_1 &= y_0 + h\,f(x_0, y_0) \\
&= 2 + 0.25(4) \\
y_1 &= 3
\end{aligned}
$$

Taking $n = 1$in Eq. (1), we have

$$
\begin{aligned}
y_2 &= y_1 + h\,f(x_1, y_1) \\
&= 3 + 0.25(3(1.25)^2 + 1) \\
&= 3 + 0.25(5.6875) \\
y_2 &= 4.4219
\end{aligned}
$$

Taking $n = 2$in Eq. (1), we have

$$
\begin{aligned}
y_3 &= y_2 + h\,f(x_2, y_2) \\
&= 4.4219 + 0.25(7.75) \\
&= 6.3594
\end{aligned}
$$

Example 2
Using Euler's method to find $y(0.1)$& $y(0.2)$given

$$y' = (x^3 + xy)e^{-x}$$

Solution
Given

$$\frac{dy}{dx} = (x^3 + xy)e^{-x}$$

Here,

$$
\begin{aligned}
x_0 &= 0, y_0 = 1, h = 0.1 \\
x_0 &= 0, x_1 = 0.1, x_2 = 0.2, x_3 = 0.3
\end{aligned}
$$

We know Euler's algorithm

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

Taking $n = 0$in Eq. (1), we have

$$y_1 = y_0 + h\,f(x_0, y_0)$$
$$= 1 + 0.1(0)$$
$$= 1$$

Also,

$$f(x_0, y_0) = (0^3 + 0(1))e^{-0} = 0$$

Taking $n = 1$in Eq. (1), we have

$$y_2 = y_1 + h\,f(x_1, y_1)$$
$$= 1 + 0.1(0.0914)$$
$$= 1.0091$$

where

$$f(x_1, y_1) = ((0.1)^3 + (0.1)(1))e^{-0.1}$$
$$= (0.101)e^{-0.1}$$
$$= 0.0914$$

Taking $n = 2$in Eq. (1), we have

$$y_3 = y_2 + h\,f(x_2, y_2)$$
$$= 1.0091 + 0.1(0.1733)$$
$$\approx 1.0264$$

where

$$f(x_2, y_2) = ((0.2)^3 + (0.2)(1.0091))e^{-0.2}$$
$$\approx 0.1733$$

Hence,

$$\boxed{y(0.1) \approx 1.0000, y(0.2) \approx 1.0091}$$

(The next iteration gives $y(0.3) \approx 1.0264$.)

Modified Euler's Theorem:

Suppose we have to solve

$$\frac{dy}{dx} = f(x, y), y(x_0) = y_0$$

to find

$$y(x_1) = y_1, \text{at } x = x_1 = x_0 + h.$$

Compute

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

where $y_1^{(n)}$ is the n^{th} approximation to y_1.
If two successive values $y_1^{(n)}$ and $y_1^{(n+1)}$ are almost equal, we stop there and take

$$y_1 = y_1^{(n)}.$$

Now we have

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

with $y = y_1 \quad x = x_1$.
To get

$$y_2 = y(x_2),$$

we use the above procedure again.

---

1. **Solve**

$$y' = y + e^x, y(0) = 0,$$

for

$$x = 0.2,\ 0.4.$$

Slution:
Given

$$f(x, y) = y + e^x, x_0 = 0, y_0 = 0,$$
$$h = 0.2.$$

To find $\boldsymbol{y_1 = y(x_1) = y(0.2)}$
Predictor:

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

Since

$$f(0,0) = 1,$$
$$y_1^{(0)} = 0 + 0.2(1) = 0.2.$$

Corrector:

$$y_1^{(1)} = y_0 + \frac{h}{2}\left[f(x_0,y_0) + f(x_1,y_1^{(0)})\right]$$
$$= 0 + \frac{0.2}{2}[\,1 + 1.4214\,]$$
$$= \frac{0.2}{2}(2.4214) = 0.2421.$$

Here,

$$f(0.2,0.2) = 0.2 + e^{0.2} = 1.4214.$$

Next correction:

$$y_1^{(2)} = y_0 + \frac{h}{2}\left[f(x_0,y_0) + f(x_1,y_1^{(1)})\right]$$
$$= 0 + \frac{0.2}{2}[\,1 + 1.4635\,]$$
$$= \frac{0.2}{2}(2.4635) = 0.2464.$$

where

$$f(0.2,0.2421) = 0.2421 + e^{0.2} = 1.4635.$$

Next,

$$y_1^{(3)} = y_0 + \frac{h}{2}\left[f(x_0,y_0) + f(x_1,y_1^{(2)})\right]$$
$$= 0 + \frac{0.2}{2}[1{+}0.2464{+}e^{0.2}].$$

=20.2(2.4678)=0.2468.

Since successive values are almost equal,

$$\boxed{y(0.2) \approx 0.2468.}$$

y1(4)=y0+2h[f(x0,y0)+f(x1,y1(3))]

$$= 0 + \frac{0.2}{2}[1{+}0.2468{+}e^{0.2}]$$
$$= 0.2468$$

Since we find $y_1^{(3)}$and $y_1^{(4)}$are equal, we take

$$y_1 = y(x_1) = y(0.2) = 0.2468.$$

Now we find y_2.
We take

$$y_1 = y(x_1) = y(0.2) = 0.2468,$$
$$x_1 = 0.2, y_1 = 0.2468, x_2 = 0.4, h = 0.2.$$

To find y_2

$$y_2^{(0)} = y(x_2) = y(0.4)$$
$$= y_1 + h\,f(x_1, y_1)$$
$$= 0.2468 + 0.2(0.2468 + 1.2214)$$
$$y_2^{(0)} = 0.5404$$

where

$$f(x_1, y_1) = 0.2468 + e^{0.2} = 1.4682.$$

$$y_2^{(1)} = y_1 + \frac{h}{2}\left[f(x_1, y_1) + f(x_2, y_2^{(0)})\right]$$
$$= 0.2468 + \frac{0.2}{2}[1.4682{+}0.5404{+}1.4918]$$
$$= 0.5968$$

$$y_2^{(2)} = y_1 + \frac{h}{2}\left[f(x_1, y_1) + f(x_2, y_2^{(1)})\right]$$
$$= 0.2468 + \frac{0.2}{2}[1.4682{+}0.5968{+}1.4918]$$
$$= 0.6035$$

$$y_2^{(3)} = y_1 + \frac{h}{2}\left[f(x_1, y_1) + f(x_2, y_2^{(2)})\right]$$
$$= 0.2468 + 0.1[1.4682{+}0.6035{+}1.4918]$$
$$= 0.6031$$

$$y_2^{(4)} = y_1 + \frac{h}{2}\left[f(x_1, y_1) + f(x_2, y_2^{(3)})\right]$$
$$= 0.2468 + \frac{0.2}{2}[1.4682{+}0.6031{+}1.4918]$$
$$= 0.6031$$

Since

$$y_2^{(3)} = y_2^{(4)} = 0.6031,$$

we have

$$\boxed{y_2 = y(0.4) = 0.6031.}$$

Using **Modified Euler's Method**, compute

$$y(0.02) \text{ and } y(0.04)$$

given

$$\frac{dy}{dx} = x^2 + y, y(0) = 1.$$

Given

$$f(x, y) = x^2 + y$$
$$x_0 = 0, y_0 = 1, h = 0.02$$

Step 1: Compute $y(0.02)$

At

$$x_0 = 0, y_0 = 1$$

First,

$$f(x_0, y_0) = 0^2 + 1 = 1$$

Predictor

$$y_1^{(0)} = 1 + 0.02(1)$$
$$= 1.02$$

First Corrector

At

$$x_1 = 0.02$$
$$f(x_1, y_1^{(0)}) = (0.02)^2 + 1.02$$
$$= 0.0004 + 1.02$$
$$= 1.0204$$

Hence,

$$y_1^{(1)} = 1 + \frac{0.02}{2}(1 + 1.0204)$$
$$= 1 + 0.01(2.0204)$$
$$= 1.020204$$

Second Corrector

$$f(0.02, 1.020204) = 0.0004 + 1.020204$$
$$= 1.020604$$

Therefore,

$$y_1^{(2)} = 1 + 0.01(1 + 1.020604)$$
$$= 1.02020604$$

Third Corrector

$$f(0.02, 1.02020604) = 1.02060604$$
$$y_1^{(3)} = 1 + 0.01(1 + 1.02060604)$$
$$= 1.02020606$$

Since the values agree up to six decimal places,

$$\boxed{y(0.02) = 1.020206}$$

Step 2: Compute $y(0.04)$
Now,

$$x_1 = 0.02, y_1 = 1.020206$$

First,

$$f(x_1, y_1) = (0.02)^2 + 1.020206$$
$$= 0.0004 + 1.020206$$
$$= 1.020606$$

Predictor

$$y_2^{(0)} = 1.020206 + 0.02(1.020606)$$
$$= 1.020206 + 0.02041212$$
$$= 1.04061812$$

First Corrector

At

$$x_2 = 0.04$$
$$f(0.04, 1.04061812) = (0.04)^2 + 1.04061812$$
$$= 0.0016 + 1.04061812$$
$$= 1.04221812$$

Hence,

$$y_2^{(1)} = 1.020206 + \frac{0.02}{2}(1.020606 + 1.04221812)$$
$$= 1.020206 + 0.01(2.06282412)$$
$$= 1.04083424$$

Second Corrector

$$f(0.04, 1.04083424) = 1.04243424$$
$$y_2^{(2)} = 1.020206 + 0.01(1.020606 + 1.04243424)$$
$$= 1.020206 + 0.02063040$$
$$= 1.04083640$$

Third Corrector

$$f(0.04, 1.04083640) = 1.04243640$$
$$y_2^{(3)} = 1.020206 + 0.01(1.020606 + 1.04243640)$$
$$= 1.04083642$$

The values have converged.
Therefore,

$$\boxed{y(0.04) = 1.040836}$$

Final Answers

$$\boxed{y(0.02) = 1.020206}$$
$$\boxed{y(0.04) = 1.040836}$$

Runge–Kutta Method:

First Order Runge–Kutta Method:

We know that by Euler's method

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

Then

$$y_1 = y_0 + h\,y_0'$$

Since

$$\frac{dy}{dx} = y' = f(x,y)$$
$$y_0' = f(x_0, y_0)$$

Expanding the left-hand side by Taylor's series, we get

$$y_1 = y(x_1) = y(x_0+h) = y_0 + hy_0' + \frac{h^2}{2!}y_0'' + \frac{h^3}{3!}y_0''' + \cdots + \frac{h^n}{n!}y_0^{(n)}$$

It follows that Euler's method agrees with Taylor's series solution up to the term in h. Hence Euler's method is the ***Runge–Kutta method of first order****.*

Second Order RK Method

We know that by the modified RK method

$$y_1^{(2)} = y_0 + \frac{h}{2}\left[f(x_0,y_0) + f(x_1,y_1^{(1)})\right]$$
$$= y_0 + \frac{h}{2}[f_0 + f(x_0+h,\ y_0+hf_0)](1)$$

where

$$hf_0 = f(x_0,y_0)$$
$$k_1 = h\,f_0$$
$$k_2 = h\,f(x_0+h,\ y_0+k_1)$$

Then Eq. (1) becomes

$$y_1 = y_0 + \frac{1}{2}(k_1+k_2)$$

which is the ***second order RK formula****.*

Hence,

2nd Order RK Formula

$$\boxed{y_1 = y_0 + \frac{1}{2}(k_1+k_2)}$$

where

$$k_1 = h\,f(x_0,y_0)$$
$$k_2 = h\,f(x_0+h,\ y_0+k_1)$$

Third Order RK Method

The 3rd order RK method is defined as

$$\boxed{y_1 = y_0 + \frac{1}{6}(k_1 + 4k_2 + k_3)}$$

where

$$k_1 = h\,f(x_0, y_0)$$
$$k_2 = h\,f\left(x_0 + \frac{h}{2}, y_0 + \frac{1}{3}k_1\right)$$
$$k_3 = h\,f(x_0 + h,\ y_0 + 2k_2 - k_1)$$

Fourth Order RK Method

The 4th order RK method is given by

$$\boxed{y_1 = y_0 + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)}$$

where

$$k_1 = h\,f(x_0, y_0)$$
$$k_2 = h\,f\left(x_0 + \frac{h}{2}, y_0 + \frac{1}{2}k_1\right)$$
$$k_3 = h\,f\left(x_0 + \frac{h}{2}, y_0 + \frac{1}{2}k_2\right)$$
$$k_4 = h\,f(x_0 + h,\ y_0 + k_3)$$

PROBLEMS

1. ***Apply RK Method of 2nd Order, find the approximate value of ywhen $x = 1.1$, given***

$$\frac{dy}{dx} = 3x + y^2, y(1) = 1.2$$

Solution

Given

$$\frac{dy}{dx} = f(x, y) = 3x + y^2$$
$$y(1) = 1.2$$
$$x_0 = 1, y_0 = 1.2$$

Here,

$$h = 0.1, x_1 = 1.1$$

1st Order RK Method

$$\begin{aligned}
&y_1 = y_0 + h\,f(x_0, y_0)\\
&= 1.2 + 0.1\,f(1,1.2)\\
&= 1.2 + 0.1[3 + (1.2)^2]\\
&= 1.2 + 0.1(4.44)\\
&y_1' = 1.644
\end{aligned}$$

2nd Order RK Method

$$\begin{aligned}
&y_1 = y_0 + \frac{1}{2}(k_1 + k_2)\\
&k_1 = h\,f(x_0, y_0)\\
&k_2 = h\,f(x_0 + h,\ y_0 + k_1)\\
&k_1 = 0.1(4.44) = 0.444\\
&k_2 = 0.1\,f(1 + 0.1,\ 1.2 + 0.444)\\
&= 0.1\,f(1.1,\ 1.644)\\
&= 0.1[3(1.1) + (1.644)^2]\\
&= 0.1(6.0027)\\
&= 0.6003
\end{aligned}$$

Hence,

$$\begin{aligned}
&y_1 = 1.2 + \frac{1}{2}(0.444 + 0.6003)\\
&= 1.2 + 0.5(1.0443)\\
&\boxed{y_1 = 1.7222}
\end{aligned}$$

---

2 . Using RK Method of Second Order, Compute $y(2.5)$ from

$$\frac{dy}{dx} = \frac{x + y}{x}, y(2) = 2$$

taking

$$h = 0.25$$

Given

$$\begin{aligned}
&\frac{dy}{dx} = f(x, y) = \frac{x + y}{x}\\
&y(2) = 2, x_0 = 2, y_0 = 2, h = 0.25, x_1 = 2.5
\end{aligned}$$

1st Order RK Method

$$y_1 = y_0 + h\,f(x_0, y_0)$$
$$= 2 + 0.25\left[\frac{2+2}{2}\right]$$
$$= 2 + 0.25(2)$$
$$y_1 = 2.5$$

2nd Order RK Method

$$y_1 = y_0 + \frac{1}{2}(k_1 + k_2)$$

where

$$k_1 = h\,f(x_0, y_0)$$
$$k_2 = h\,f(x_0 + h,\ y_0 + k_1)$$

Now,

$$k_1 = 0.25(2) = 0.5$$
$$k_2 = 0.25\,f(2.25,\ 2.5)$$
$$= 0.25\left[\frac{2.25 + 2.5}{2.25}\right]$$
$$= 0.25(2.1111)$$
$$k_2 = 0.5278$$

Hence,

$$y_1 = 2 + \frac{1}{2}(0.5 + 0.5278)$$
$$= 2 + 0.5(1.0278)$$
$$\boxed{y_1 = 2.5139}$$

Runge–Kutta Method (Fourth Order) – Solved Problem

1. **Problem**

Using the **Runge–Kutta Fourth Order (RK4) Method**, find the value of $y$at $x = 0.2$given that

$$\frac{dy}{dx} = x + y, y(0) = 1,$$

taking step size

$$h = 0.2.$$

Formula (RK4)

$$k_1 = h\,f(x_n, y_n)$$
$$k_2 = h\,f\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right)$$
$$k_3 = h\,f\left(x_n + \frac{h}{2}, y_n + \frac{k_2}{2}\right)$$
$$k_4 = h\,f(x_n + h, y_n + k_3)$$

Then,

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4).$$

Given $f(x, y) = x + y$

$$x_0 = 0, y_0 = 1, h = 0.2$$

Step 1: Calculate k_1

$$k_1 = 0.2(0 + 1)$$
$$k_1 = 0.2$$

Step 2: Calculate k_2

$$k_2 = 0.2\left[0.1 + \left(1 + \frac{0.2}{2}\right)\right]$$
$$= 0.2(0.1 + 1.1)$$
$$= 0.2(1.2)$$
$$k_2 = 0.24$$

Step 3: Calculate k_3

$$k_3 = 0.2\left[0.1 + \left(1 + \frac{0.24}{2}\right)\right]$$
$$= 0.2(0.1 + 1.12)$$
$$= 0.2(1.22)$$
$$k_3 = 0.244$$

Step 4: Calculate k_4

$$k_4 = 0.2[0.2 + (1 + 0.244)]$$
$$= 0.2(1.444)$$
$$k_4 = 0.2888$$

Step 5: Calculate y_1

$$y_1 = 1 + \frac{1}{6}(0.2 + 2(0.24) + 2(0.244) + 0.2888)$$
$$= 1 + \frac{1}{6}(1.4568)$$
$$= 1 + 0.2428$$
$$\boxed{y(0.2) \approx 1.2428}$$

Final Answer

$$\boxed{y(0.2) = 1.2428}$$

PRACTICE QUESTIONS:

1. Solve $\frac{dy}{dx} = x + y,\ y(0) = 1$and find $y(0.1)$using Taylor's series method.
2. Solve $\frac{dy}{dx} = x^2 + y,\ y(0) = 1$and compute $y(0.2)$
3. Solve $\frac{dy}{dx} = x + y,\ y(0) = 1$and find $y(0.2)$using Euler's method with $h = 0.1$.
4. Solve $\frac{dy}{dx} = x - y,\ y(0) = 1$and find $y(0.3)$using Euler's method with $h = 0.1$.
5. Solve $\frac{dy}{dx} = y - x,\ y(0) = 1$and compute $y(0.2)$using the Modified Euler's method.
6. Solve $\frac{dy}{dx} = x + y^2,\ y(0) = 1$and find $y(0.2)$
7. Solve $\frac{dy}{dx} = x + y,\ y(0) = 1$and determine $y(0.2)$using RK2 with $h = 0.1$.
8. Solve $\frac{dy}{dx} = y - x^2,\ y(0) = 1$and find $y(0.2)$using RK4 with $h = 0.1$.
9. Solve $\frac{dy}{dx} = x + y,\ y(1) = 2$and compute $y(1.2)$using RK4 with $h = 0.1$.
10. Solve $\frac{dy}{dx} = x + y,\ y(0) = 1$using Picard's method and obtain the first three approximations

UNIT-4

Estimation and Testing of Hypothesis – Large Sample Tests

UNIT-4

Estimation and Testing of Hypothesis – Large Sample Test

Introduction

Statistics plays a vital role in engineering, science, medicine, business, and research by enabling decisions to be made based on data rather than assumptions. In practical situations, engineers and researchers often need to verify claims about a population using only a sample of observations. This process is known as **hypothesis testing**.

A **statistical hypothesis** is an assumption or statement about a population parameter, such as the population mean, proportion, or variance. Since it is usually impractical to study an entire population, a random sample is selected, and statistical techniques are used to determine whether the available evidence supports or rejects the stated hypothesis.

Large sample tests are applicable when the sample size is **30 or more ($n \geq 30$)**. These tests are based on the **Normal (Z) Distribution** and provide reliable statistical inference for testing population means and proportions. The most commonly used large sample tests include the **Z-test for a single mean, Z-test for the difference between two means, Z-test for a single proportion, and Z-test for the difference between two proportions**. Confidence intervals and confidence limits are also important tools for estimating population parameters.

Hypothesis testing is widely applied in engineering for **quality control, product reliability analysis, process improvement, performance evaluation, manufacturing, healthcare, agriculture, and scientific research**. It helps engineers and researchers make objective, data-driven decisions with a specified level of confidence.

This unit provides the fundamental concepts of hypothesis testing, introduces the principles of large sample tests, and explains their applications through suitable mathematical formulations, solved examples, and engineering case studies.

Learning Objectives

After studying this unit, students will be able to:

- Understand the concept of statistical hypothesis testing.
- Distinguish between null and alternative hypotheses.
- Explain the concepts of level of significance, critical region, and decision making.
- Perform large sample **Z-tests** for population means and proportions.
- Construct confidence intervals and confidence limits.
- Interpret statistical results in engineering and scientific applications.
- Apply hypothesis testing techniques to solve real-world engineering problems.

Real-Life Engineering Applications

- Quality control in manufacturing industries.
- Performance testing of machines and equipment.
- Reliability analysis of electronic components.
- Comparison of production processes.
- Clinical trials and medical research.
- Agricultural yield analysis.
- Market research and customer satisfaction surveys.

Prerequisites

Before studying this unit, students should be familiar with:

- Measures of Central Tendency
- Measures of Dispersion
- Probability Distributions
- Normal Distribution
- Sampling Techniques
- Standard Error

Outcomes

On successful completion of this unit, students will be able to:

1. Formulate statistical hypotheses for engineering problems.
2. Select appropriate large sample statistical tests.
3. Apply Z-tests for testing means and proportions.
4. Construct and interpret confidence intervals.
5. Draw valid conclusions based on statistical evidence for engineering decision-making.

Definition

1. Hypothesis

A **hypothesis** is a tentative statement or assumption about a population parameter that is tested statistically using sample data.

Example:

"The average life of a battery is 10 hours."

---

2. Statistical Hypothesis

A **statistical hypothesis** is a statement regarding one or more population parameters, such as the population mean, proportion, or variance, which is verified using statistical methods

3. Null Hypothesis (H_0)

The **Null Hypothesis (H_0)** is the hypothesis that assumes **no significant difference** or **no effect** exists. It represents the status quo and is tested for possible rejection.

OR

A hypothes is to be tested under the investigation. It is denoted by $\mathbf{H_0}$.

Ex: $H_0 : \mu = \mu_0$

Example:

$$H_0:\ \mu = 50$$

This means the population mean is **50**.

---

4. Alternative Hypothesis (H_1 or H_a)

The **Alternative Hypothesis (H_1)** is the hypothesis that contradicts the null hypothesis. It indicates that a significant difference or effect exists.

OR

Any hypothesis it is complementary to null hypothesis. It is denoted by **H_1**.

Examples:

$$H_1:\ \mu \neq 50$$
$$H_1:\ \mu > 50$$
$$H_1:\ \mu < 50$$

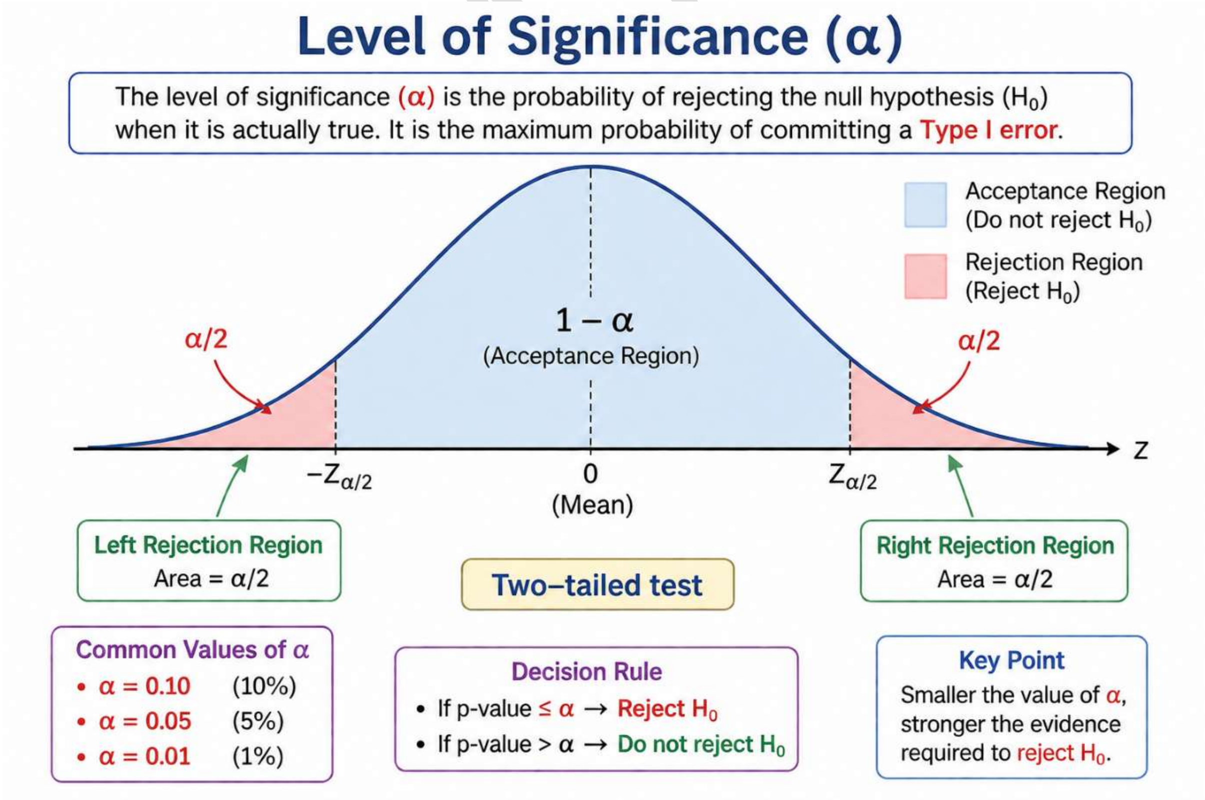
---

5. Level of Significance (α)

The **Level of Significance** is the maximum probability of rejecting a true null hypothesis. Common significance levels are **0.05 (5%)** and **0.01 (1%)**.

OR

The significant level indicates the confidence with which an experimental rejects a null hypothesis. It is usually expressed as a %, say **5%, 1%, 2%, 10%**.



6. Test Statistic

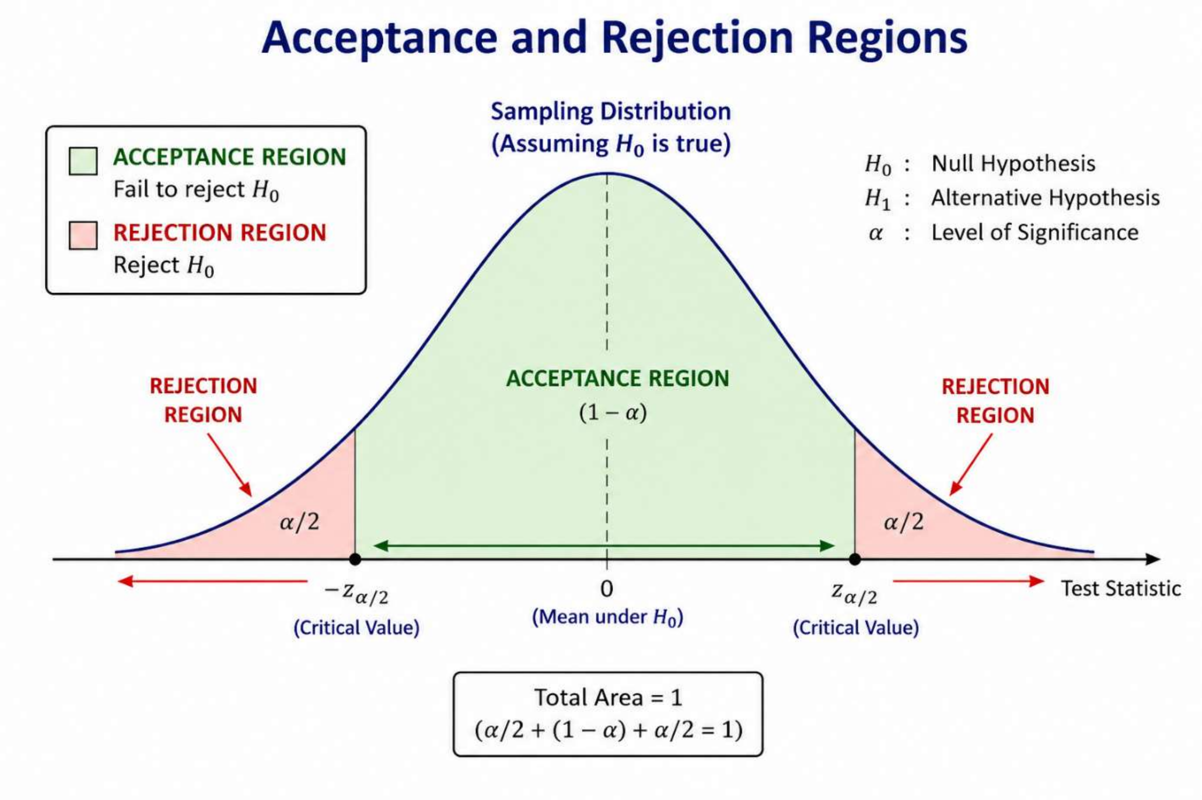
A **Test Statistic** is a numerical value calculated from the sample data and used to decide whether the null hypothesis should be accepted (fail to reject) or rejected.
For large samples, the **Z-statistic** is used.

7. Critical Region

The **Critical Region** is the region of values of the test statistic that leads to the rejection of the null hypothesis.

8. Acceptance Region

The **Acceptance Region** is the region in which the calculated test statistic supports the null hypothesis (fail to reject H_0).



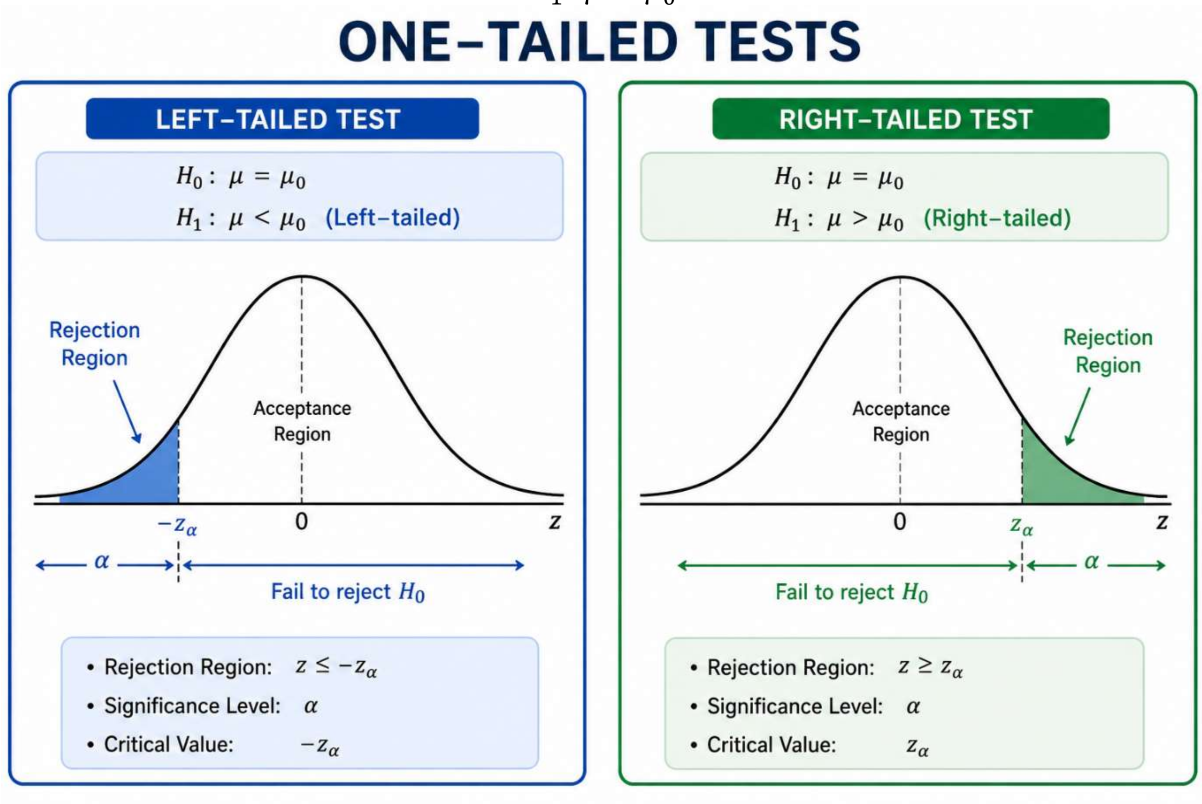
9. One-Tailed Test

A **One-Tailed Test** is used when the alternative hypothesis specifies a direction (greater than or less than).
Examples:
Right-tailed:

$$H_1: \mu > \mu_0$$

Left-tailed:

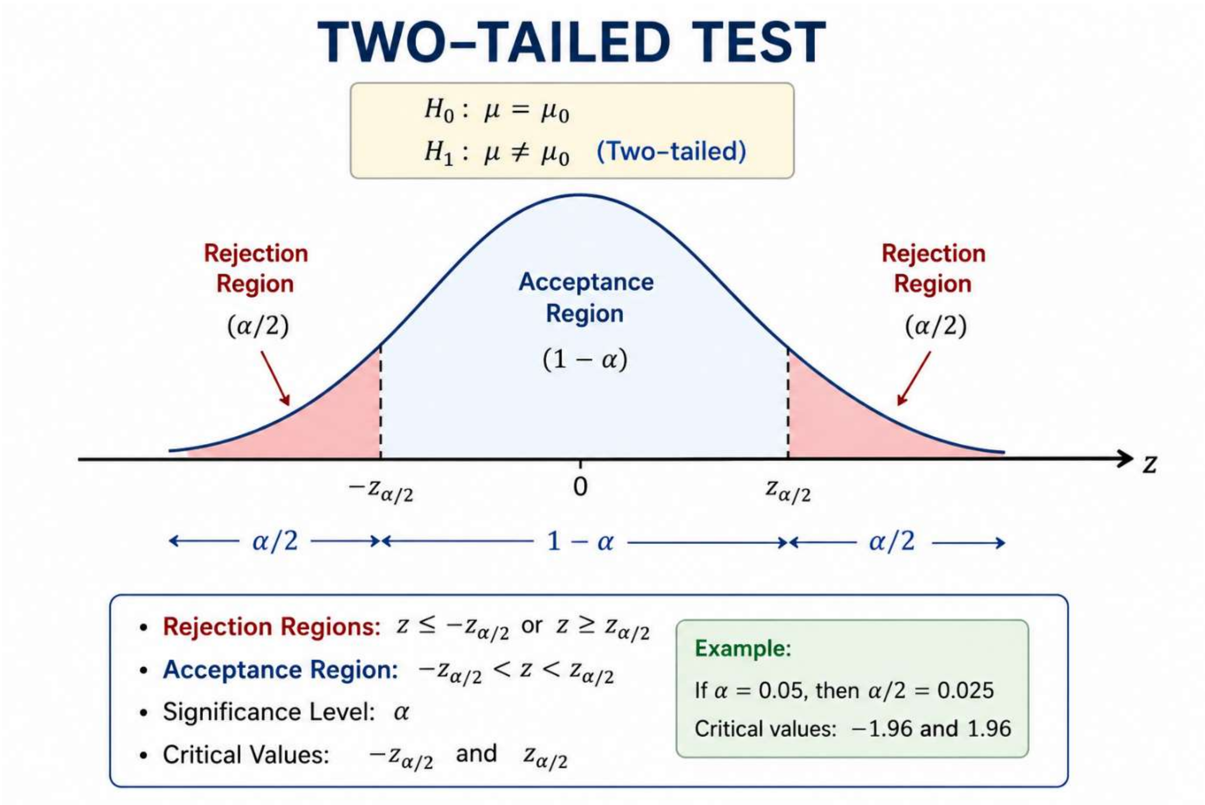
$$H_1: \mu < \mu_0$$



10. Two-Tailed Test

A **Two-Tailed Test** is used when the alternative hypothesis indicates that the parameter is simply different.

$$H_1: \mu \neq \mu_0$$



11. Errors in Sampling:

To accept or reject the hypothesis, gives the same errors or risks.
They are two types of errors or risks in the testing of hypothesis. They are:

1. Type-I error
2. Type-II Error.

1.Type I Error

A **Type I Error** occurs when a true null hypothesis is rejected.
Probability of Type I Error = **α**

2. Type II Error

A **Type II Error** occurs when a false null hypothesis is accepted (fail to reject).
Probability of Type II Error = **β**

Mathematical Formulae

1. Z-Test for Single Mean

When population standard deviation (σ) is known and sample size **n ≥ 30**,

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

Where:

- $\bar{x}$= Sample Mean
- μ= Population Mean
- σ= Population Standard Deviation
- n= Sample Size

---

2. Z-Test for Difference Between Two Means

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Where:

- $\bar{x}_1,\ \bar{x}_2$= Sample Means
- $\sigma_1,\ \sigma_2$= Population Standard Deviations
- $n_1,\ n_2$= Sample Sizes

---

3. Z-Test for Single Proportion

$$Z = \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}}$$

Where:

- p= Sample Proportion
- P= Population Proportion
- n= Sample Size

---

4. Z-Test for Difference Between Two Proportions

First calculate the pooled proportion:

$$P = \frac{x_1 + x_2}{n_1 + n_2}$$
$$Q = 1 - P$$

Then,

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Where:

- p_1, p_2= Sample Proportions
- P= Pooled Proportion
- $Q = 1 - P$

---

5. Confidence Interval for Mean

$$\bar{x} \pm Z\left(\frac{\sigma}{\sqrt{n}}\right)$$

---

6. Confidence Interval for Proportion

$$p \pm Z\sqrt{\frac{p(1-p)}{n}}$$

Standard Critical Values

Confidence Level	Significance Level (α)	Critical Z-Value
90%	0.10	1.645
95%	0.05	1.96
99%	0.01	2.576

TESTING OF HYPOTHESIS – LARGE SAMPLE TEST

PROCEDURE:

The following steps are followed to perform a **Large Sample (Z-Test)**.

Step 1: State the Hypotheses

Formulate the hypotheses based on the given problem.

- **Null Hypothesis (H_0):** Assumes no significant difference exists.
- **Alternative Hypothesis (H_1):** Assumes a significant difference exists.

Example:

$$H_0: \mu = 50$$
$$H_1: \mu \neq 50$$

Step 2: Choose the Level of Significance (α)

Select the significance level.

Common values are:

- 5% (α = 0.05)
- 1% (α = 0.01)

Step 3: Select the Appropriate Test

If sample size **n ≥ 30**, use the **Z-Test**.

Identify whether the problem involves:

- Single Mean
- Difference Between Two Means
- Single Proportion
- Difference Between Two Proportions

Step 4: Calculate the Test Statistic
Use the appropriate Z-test formula.

Step 5: Determine the Critical Value
From the Standard Normal Distribution Table:

- Two-tailed test (5%) → ±1.96
- One-tailed test (5%) → ±1.645
- Two-tailed test (1%) → ±2.576

Step 6: Make the Decision

- If the calculated Z-value lies within the acceptance region, **Accept (Fail to Reject) H_0**.
- If the calculated Z-value lies in the critical region, **Reject H_0**.

Step 7: Write the Conclusion
Interpret the statistical result in words related to the problem.

Test of Significance for Large Samples

Under large samples we see four important tests of significance. They are

1. Testing of significance for single mean.
2. Testing of significance for difference of means.
3. Testing of significance for single proportion.
4. Testing of significance for difference of proportions.

1. Testing of Significance for Single Mean

- If the population variance σ^2 is known then to test the hypothesis we use the test statistic

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

where

$$\bar{x} = \text{Sample mean}$$
$$\mu = \text{Population mean}$$
$$\sigma = \text{Population S.D}$$
$$n = \text{Sample size}$$

Here

$$\frac{\sigma}{\sqrt{n}}$$

is the standard error of sample mean $\bar{x}$.

- If the population variance σ^2 is unknown then to test the hypothesis we use the test statistic

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

where S is sample S.D.
The confidence limits (fiducial limits) for single mean at the confidence are

$$\left(\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

where $Z_{\alpha/2}$is the value of Z at the % of confidence.

PROBLEMS (Single Mean)

Q1)

A company claims that the average weight of a packet is **500 g**.
A random sample of **100 packets** has:

- Sample Mean = **495 g**
- Population Standard Deviation = **20 g**

Test the claim at the **5% level of significance**.

Solution

Given

$$n = 100, \bar{x} = 495, \mu = 500, \sigma = 20$$

Step 1: Hypotheses

$$H_0: \mu = 500$$
$$H_1: \mu \neq 500$$

Step 2: Formula

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

Step 3: Calculation

$$Z = \frac{495 - 500}{20/\sqrt{100}}$$
$$= \frac{-5}{2}$$
$$Z = -2.5$$

Step 4: Critical Value

At 5% significance,

$$Z = \pm 1.96$$

Step 5: Decision

Since

$$|-2.5| > 1.96$$

Reject H₀.

Conclusion

There is sufficient evidence to conclude that the average packet weight is **not 500 g**.

Q2)

In a random sample of **60 workers**, the average time taken by them to get to work is **33.8 min** with a S.D of **6.1 min**. Can we reject the null hypothesis **μ = 32.6 min** in favour of alternative hypothesis **μ > 32.6** at **α = 0.025 L.O.S**

SOLUTION:

Given

$$n = 60$$
$$\bar{x} = 33.8$$
$$\mu = 32.6$$
$$\sigma = 6.1$$

1. **Null hypothesis:** We state the null hypothesis as
$$H_0: \mu = 32.6$$

2. **Alternative Hypothesis:**
$$H_1: \mu > 32.6$$

3. **L.O.S**
$$\alpha = 0.025, Z_{cri} = 2.33$$

4. **Test statistic:** To test the above null hypothesis, we use the test statistic

$$Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{33.8 - 32.6}{6.1/\sqrt{60}} = 1.5238$$

5. **Conclusion:**

$$Z_{cal} < Z_{cri}$$

we accept the null hypothesis.
∴ Our null hypothesis is

$$H_0: \mu = 32.6$$

Q3)
A sample of **64 students** have a mean weight of **70 kgs**. Can this be regarded as a sample from a population with mean weight **56 kgs** & S.D **25 kgs**.

Sol:

Sample

$$n = 64 \ (> 30)$$

It is a large sample.

$$\bar{x} = 70 \ kgs$$

Population

$$\mu = 56$$
$$\sigma = 25 \ kgs$$

(i) Null Hypothesis:
write given question full (sample is taken from the population with mean = 56)
i.e.

$$H_0: \mu = 56 \ kgs$$

(ii) Alternative Hypothesis:
Sample cannot be regarded as one coming from the population.
i.e.

$$H_1: \mu \neq 56 \ kgs$$

(Two tailed test)
(iii) L.O.S

$$\alpha = 0.05$$
$$(Z_{cri} = 1.96)$$

(iv) Test Statistic:

$$Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{70 - 56}{25/\sqrt{64}} = 4.48$$

(v) Conclusion:

$$|\, Z_{cal} \,| > Z_{cri}$$

value, we reject H_0.
∴ Sample cannot be regarded as one coming from the population.

Q4)

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

Sol:

Given that

Sample size

$$n = 36 > 30$$

It is a large sample.
Sample mean

$$\bar{x} = 11 \text{ min}$$

Population mean

$$\mu = 10$$

Population variance

$$\sigma^2 = 16$$

Population S.D

$$\sigma = \sqrt{16} = 4$$

(i) Null Hypothesis

$$H_0 : \mu = 10$$

(ii) Alternative Hypothesis

$$H_1 : \mu < 10$$

(Left one-tailed test)
(iii) Level of significance

$$\alpha = 0.05, Z_{cri} = 1.645$$

(iv) Test Statistic

$$Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{11 - 10}{4/\sqrt{36}} = \frac{6}{4} = 1.5$$

(v) Conclusion

Since

$$|Z_{cal}| < Z_{cri}$$

we accept H_0.

Q5)

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

Solution:

Given

$$n = 400, \bar{x} = 40, \mu = 38, \sigma = 10$$

(i) Null Hypothesis

$$H_0: \mu = 38$$

(ii) Alternative Hypothesis

$$H_1: \mu \neq 38$$

(Two tailed test)

(iii) Level of significance

$$\alpha = 0.05, Z_{cri} = 1.96$$

(iv) Test Statistic

$$Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{40 - 38}{10/\sqrt{400}} = \frac{2}{0.5} = 4$$

Conclusion

Since

$$|Z_{cal}| > Z_{cri}$$

we reject H_0.

i.e. the sample is **not** from the population whose mean is **38**.

95% Confidence interval

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

i.e.

$$\left(40 - \frac{1.96(10)}{\sqrt{400}},\ 40 + \frac{1.96(10)}{\sqrt{400}}\right)$$
$$= (40 - 0.98,\ 40 + 0.98)$$
$$= (39.02,\ 40.98)$$

Q6) A die is tossed 960 times. If 5 falls upwards 184 times, is the die unbiased at 0.01 L.O.S.

Sol:

$$n = 960$$

Probability of throwing **5** with one die

$$p = \frac{1}{6}$$
$$q = 1 - \frac{1}{6} = \frac{5}{6}$$
$$\mu = np = 960\left(\frac{1}{6}\right) = 160$$
$$\sigma = \sqrt{npq} = \sqrt{960\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)} = 11.55$$

No. of Success

$$x = 184$$

(i) Null Hypothesis

H_0:The die is unbiased.

(ii) Alternative Hypothesis

H_1:The die is biased.

(iii) L.O.S

$$\alpha = 0.01$$

(iv) Test Statistic

To test the above hypothesis we use the test statistic

$$Z = \frac{x-\mu}{\sigma} = \frac{184-160}{11.55} = 2.07$$

Z_{cri} value at 1% L.O.S is 2.58

Since

$$Z_{cal} < Z_{cri}$$

we accept H_0 at **1% L.O.S**.

Conclusion:

The die is **unbiased**.

Q6)

The mean life-time of a sample of 100 bulbs produced by a company is found to be 1560 hours with a population S.D = 90 hours. Test the hypothesis that the mean life-time of the bulbs produced by the company is 1580 hours at 5% and 1% level of significance.

Solution:

Given that

- $n = 100 > 30 \Rightarrow$ It is a large sample.
- $\bar{x} = 1560\text{hr}$
- $\sigma = 90\text{hr}$

1. **Null hypothesis (H_0):**
   Let $H_0: \mu = 1580\text{hr}$
2. **Alternative hypothesis (H_1):**
   Let $H_1: \mu \neq 1580\text{hr}$ *(Two-tailed test)*
3. **Significance level:**

(i) $\alpha = 5\%$

(ii) $\alpha = 1\%$

4. **Test Statistic:**

$$Z_{cal} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{1560-1580}{90/\sqrt{100}} = \frac{-20}{9} = -2.22$$

$$|Z_{cal}| = 2.22$$

5. **Conclusion:**

(i) At $\alpha = 5\%$:

$$Z_{cri} = 1.96$$

Since

$$|Z_{cal}| > Z_{cri}$$

Reject H_0.

$$\therefore \mu \neq 1580 \text{ hr}$$

(ii) At $\alpha = 1\%$:

$$Z_{cri} = 2.58$$

Since

$$|Z_{cal}| < Z_{cri}$$

Accept H_0.

$$\therefore \mu = 1580 \text{ hr}$$

Test of Hypothesis for Two Means:

Given

$$\mu_1 - \mu_2 = 0$$
$$\mu_1 = \mu_2$$

Null hypothesis:

$$H_0: \mu_1 = \mu_2$$

Alternative hypothesis:

$$H_1: \mu_1 \neq \mu_2$$

(or)

$$\mu_1 < \mu_2$$

(or)

$$\mu_1 > \mu_2$$

Test Statistic

If σ_1, σ_2 are **known**,

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

If σ_1, σ_2 are **unknown**,

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

For testing

$$\mu_1 - \mu_2 = d$$

Null hypothesis:

$$H_0: \mu_1 - \mu_2 = d$$

Alternative hypothesis:

$$H_1: \mu_1 - \mu_2 \neq d$$

or

$$\mu_1 - \mu_2 < d$$

or

$$\mu_1 - \mu_2 > d$$

The test statistic is

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

where

- $\bar{x}_1$= first sample mean
- $\bar{x}_2$= second sample mean
- n_1= first sample size
- n_2= second sample size
- σ_1= first population S.D.
- σ_2= second population S.D.
- S_1= first sample S.D.
- S_2= second sample S.D.

Problems:

Question1:

Two factories produce steel rods.

Factory	Mean Length (cm)	Standard Deviation	Sample Size
A	102	8	64
B	100	6	49

Test whether the average lengths differ significantly at the **5% level of significance**.

Solution

Given

$$\bar{x}_1 = 102, \bar{x}_2 = 100$$
$$\sigma_1 = 8, \sigma_2 = 6$$
$$n_1 = 64, n_2 = 49$$

Formula

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Calculation

$$Z = \frac{102 - 100}{\sqrt{\frac{64}{64} + \frac{36}{49}}}$$
$$= \frac{2}{1.317}$$
$$Z = 1.52$$

Critical value

$$Z = \pm 1.96$$

Since

$$1.52 < 1.96$$

Accept (Fail to Reject) H_0.

Conclusion

There is **no significant difference** between the average lengths produced by the two factories.

Question2:

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 the same population of **S.D. = 2.5 inches**?

Solution:

Given

- $n_1 = 1000,\ n_2 = 2000$
- $\bar{x}_1 = 67.5,\ \bar{x}_2 = 68$
- $\sigma = 2.5$

1. **Null hypothesis**
   H_0:The samples have been drawn from the same population of S.D. 2.5 inches.

2. **Alternative hypothesis**
   H_1:The samples are not drawn from the same population of S.D. 2.5 inches.

3. **Level of Significance**

$$\alpha = 0.05, Z_{cri} = 1.96$$

4. **Test Statistic**

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

$$= \frac{67.5 - 68}{\sqrt{\frac{(2.5)^2}{1000} + \frac{(2.5)^2}{2000}}}$$

$$= \frac{-0.5}{0.0968} = -5.16$$

Hence,

$$|\ Z_{cal}\ | = 5.16$$

5. **Conclusion**

$$Z_{cri} = 1.96$$

Since

$$|\ Z_{cal}\ | > Z_{cri}$$

Reject H_0.

∴The samples are **not drawn from the same population** of S.D. **2.5 inches**.

Question3:

A sample of height of **6400 Englishmen** has a mean **67.85 inches** and S.D of **2.56 inches**, while a sample of heights of **1600 Australians** has a mean of **68.55 inches** and S.D of **2.52 inches**. Do the data indicate Australians are on the average **taller** than the English? *(use α = 0.01)*

Solution:

Given

$$n_1 = 6400, \bar{x}_1 = 67.85, s_1 = 2.56 \rightarrow \text{English}$$
$$n_2 = 1600, \bar{x}_2 = 68.55, s_2 = 2.52 \rightarrow \text{Australians}$$

1. **Null hypothesis** : $H_0: \mu_1 = \mu_2$
2. **Alternative Hyp. (H_1)** : $\mu_1 < \mu_2$
3. **Level of Significance** :

$$\alpha = 0.01 = 1\%$$

4. **Test Statistic** :

$$|Z| = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = |9.9| = 9.9$$

5. **Conclusion** :

By Table,

$$Z_{cri} = +2.33$$
$$|Z_{cal}| > |Z_{cri}|$$

$\therefore H_0$ is rejected.

i.e. Thus Australians are taller than Englishmen.

Question4:

It has been claimed that the **resistance of electric wire** can be reduced by **more than 0.050 ohm** by alloying. To test this claim, **32 values** obtained for standard wire yielded $\bar{x}_1 = 0.136$ ohm, $s_1 = 0.004$ ohm and **32 values** obtained for alloyed wire yielded $\bar{x}_2 = 0.083$ ohm, $s_2 = 0.005$ ohm Does this sample support the claim? At **0.050** L.O.S

Solution:

Given

$$n_1 = 32, \bar{x}_1 = 0.136, \bar{x}_2 = 0.083, s_1 = 0.004, s_2 = 0.005, \alpha = 5\%$$

1. **Null Hypothesis** :

$$H_0: \mu_1 - \mu_2 = 0.050$$

2. **Alternative Hypothesis** :

$$H_1: \mu_1 - \mu_2 > 0.050$$

3. **L.O.S** :

$$\alpha = 5\%, Z_{cri} = 1.645$$

4. **Test Statistic** :

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\mid Z_{cal} \mid = \frac{(0.136 - 0.083) - 0.050}{\sqrt{\frac{(0.004)^2}{32} + \frac{(0.005)^2}{32}}} = 2.65$$

Conclusion :

$$\mid Z_{cal} \mid > \mid Z_{cri} \mid$$

$\therefore H_0$ is rejected.

Test of Hypothesis for One Proportion (Without mean & SD given)

The test statistic for test of hypothesis for one proportion is given by

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

where

- p= Sample proportion
- P= Population proportion
- $Q = 1 - P$
- n= Sample size

Confidence Interval (or) Confidence Limits:

Confidence interval of population proportion P is

$$\left(p - Z_{\alpha/2}\sqrt{\frac{PQ}{n}},\ p + Z_{\alpha/2}\sqrt{\frac{PQ}{n}}\right)$$

Procedure:

1. $H_0: P = 0.5$
2. $H_1: P \neq 0.5$ *(Two-tail)*
   $P < 0.5$ or $P > 0.5$ *(One tail)*
3. $\alpha = 5\%$ or 1% or 2% or 10%
4. Test Statistic

$$Z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

Z Table

α	Two tail	One tail
5%	1.96	1.645
1%	2.58	2.33
2%	2.33	—
10%	1.645	—

5. **Conclusion**

If

$$| Z_{cal} | > | Z_{cri} |$$

Reject H_0.

Examples:

1. In a big city **325 men out of 600 men** were found to be smokers. Does the information support the conclusion that the majority of men in this city are smokers?

Solution:

$$n = 600, p = \frac{325}{600} = 0.5417$$

Let

$$P = \frac{1}{2}$$
$$Q = 1 - P = 1 - \frac{1}{2} = \frac{1}{2}$$

1. **Null Hypothesis**

$$H_0: P = 0.5$$

(i.e. the no. of smokers & non-smokers are equal in the city)

2. **Alternative Hypothesis**

$$H_1: P > 0.5$$

3. **L.O.S**

$$\alpha = 5\%$$
$$| Z_{cri} | = 1.645$$

4. **Test Statistic**

$$Z=\frac{p-P}{\sqrt{\frac{PQ}{n}}}=\frac{0.5417-0.5}{\sqrt{\frac{0.5\times0.5}{600}}}=2.04$$

5. **Conclusion**

Since

$$\mid Z_{cal}\mid>\mid Z_{cri}\mid$$

$H_0$is rejected.

i.e. the majority of men in the city are **smokers**.

2. In a big city 250 men out of 750 men were found to be smokers. Does the information support the conclusion that the majority of men in this city are smokers?

$$n=750, p=\frac{250}{750}=0.333$$

$$P=\frac{1}{2}, Q=1-P=1-\frac{1}{2}=\frac{1}{2}$$

1. **Null Hypothesis:**

$$H_0\colon P=0.5$$

2. H_1**:**

$$P>0.5$$

3. **L.O.S:**

$$\alpha=5\%, \mid Z_{cri}\mid=1.645$$

4. **Test Statistic:**

$$Z_{cal}=\frac{p-P}{\sqrt{\frac{PQ}{n}}}=\frac{0.333-0.5}{\sqrt{\frac{0.5\times0.5}{750}}}=\frac{-0.167}{0.01826}\approx-9.15$$

5. **Conclusion**

$$\mid Z_{cri}\mid>\mid Z_{cal}\mid$$

So we reject the claim that majority are smokers.

3.A die was thrown **9000 times** and of these **3000 yielded 3 or 4**. Is this consistent with the hypothesis that the die was unbiased? *(use* $\alpha=0.01$*)*

Given

$$n=9000, \text{Sample proportion } p=\frac{3000}{9000}=0.3333$$

Population proportion of getting **3 or 4**

$$P = \frac{2}{6} = \frac{1}{3} = 0.3333$$
$$Q = 1 - P = 1 - 0.3333 = 0.6667$$

1. **Null Hypothesis:**

$$H_0: P = \frac{1}{3}$$

(i.e. the die is unbiased)

2. **Alternative Hypothesis:**

$$H_1: P \neq \frac{1}{3}$$

(The die is biased)

3. **L.O.S**

$$\alpha = 1\%$$

4. **Test Statistic:**

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.3578 - 0.3333}{\sqrt{\frac{0.3333 \times 0.6667}{9000}}} = 4.94$$

5. **Conclusion:**

$$|Z_{cal}| > |Z_{cri}|$$

$H_0$is rejected, i.e. **The die is biased.**

---

4. A manufacturer claimed that atleast 95% of the equipment which be supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test this claim at 1% L.O.S.

Given

$n = 200$,No. of pieces conforming to specifications $= 200 - 18 = 182$

$p =$ proportion of pieces conforming to specifications $= \frac{182}{200} = 0.91$

P = population proportion = 95% = 0.95

$$Q = 1 - P = 0.05$$

1. $H_0: P = 95\%$
2. *A.H :* $H_1: P > 0.95$

3. *L.O.S :* $\alpha = 5\%$
4. *Test statistic :*

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = -2.59$$

5. ***Conclusion :***

$$Z_{cri} = 1.645, |\, Z_{cal} \,| = 2.59$$
$$|\, Z_{cal} \,| > |\, Z_{cri} \,| \therefore H_0 \text{ is rejected.}$$

Hence the manufacture claim is accepted.

---

5. ***If 80 patients are treated with an antibiotic 59 got cared. Find a 99% confidence limits to the true population.***

Given

$$n = 80, p = \frac{59}{80} = 0.7375$$
$$q = 1 - p = 1 - 0.7375 = 0.2625$$

*Confidence limit = **99%***

$Z_{\frac{\alpha}{2}}$ *for 99% Confidence =2.58*

Confidence interval for P *is*

$$\left(p - Z_{\alpha/2}\sqrt{\frac{pq}{n}},\ p + Z_{\alpha/2}\sqrt{\frac{pq}{n}}\right)$$
$$= (0.7375 - 2.58 \times 0.049,\ 0.7375 + 2.58 \times 0.049)$$
$$= (0.61,\ 0.86)$$

---

6. ***In a random sample of 400 industrial accidents, it was found that 231 were due to unsafe working conditions. What can we say at 95% confidence about the maximum error.***

Given

$$p = \frac{231}{400} = 0.58, q = 1 - p = 0.42, n = 400$$

$Z_{\frac{\alpha}{2}}$ *for 95% Confidence =1.96*

Maximum error

$$E = Z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

$$= 1.96\sqrt{\frac{0.58 \times 0.42}{400}}$$
$$= 0.048$$

Test of Hypothesis of Two Proportions

Test statistic

$$Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}, Q = 1 - P$$

n_1= *first sample size*
n_2= *second sample size*
P_1= *first sample proportion*
P_2= *second sample proportion*

Note:

Suppose the population P_1*and* P_2*are given and* $P_1 \neq P_2$.
If we want to test the Hypothesis that the difference $(P_1 - P_2)$*in population proportion is likely to be hidden in samples of sizes* n_1 & n_2*from two populations respectively.*

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

If sample proportions are not known then we use

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

PROCEDURE:

1.$H_0: P_1 = P_2$
2.$H_1: P_1 \neq P_2$*(Two-tail)*
$P_1 > P_2$*(or)* $P_1 < P_2$*(one tail)*
3.α: 5% *(or)* 1% *or* 2% *or* 10%.

4.Test statistic

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$$P_1 = \frac{x_1}{n_1}, P_2 = \frac{x_2}{n_2}$$
$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$
$$q = 1 - p$$

5.Conclusion:

$$|\, Z_{cal} \,| \leq Z_{cri}, \text{Accept } H_0.$$

PROBLEMS:

1) *Two large populations there are **30% & 25% respectively** for fair haired people. Is this difference likely to be hidden in sample of **1200 & 900** respectively from two population.*

SOLUTION:

Given

$$n_1 = 1200, n_2 = 900$$

1. H_0: $P_1 = P_2$
2. H_1: $P_1 \neq P_2$ *(two tail.)*
3. *L.O.S:*

$$\alpha = 5\%, |\, Z_{cri} \,| = 1.96$$

4. *Test statistic:*

$$n_1 = 1200, n_2 = 900$$
$$P_1 = 30\% = \frac{30}{100} = 0.3, P_2 = 25\% = \frac{25}{100} = 0.25$$
$$Q_1 = 1 - P_1 = 1 - 0.3 = 0.7$$
$$Q_2 = 1 - P_2 = 1 - 0.25 = 0.75$$
$$Z_{cal} = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{1200} + \frac{P_2 Q_2}{900}}} = \frac{0.3 - 0.25}{\sqrt{\frac{0.3 \times 0.7}{1200} + \frac{0.25 \times 0.75}{900}}} = 2.85$$

5.Conclusion:

$$|\, Z_{cal} \,| > Z_{cri}$$

Reject H_0.
i.e. $P_1 \neq P_2$.

2.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 were in favour of the proposal. Test the hypothesis that proportions of men and women in favour of the proposal are same at 5% level.

SOLUTION:

$$n_1 = 400, x_1 = 200$$
$$n_2 = 600, x_2 = 325$$
$$P_1 = \frac{x_1}{n_1} = \frac{200}{400} = 0.5, P_2 = \frac{325}{600} = 0.541$$
$$P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{200 + 325}{400 + 600} = 0.525$$
$$Q = 1 - P = 1 - 0.525 = 0.475$$

1. ***Null Hypothesis:*** $H_0: P_1 = P_2$
2. ***Alternative Hypo:*** $H_1: P_1 \neq P_2$*(two tailed test)*
3. ***L.O.S:*** $\alpha = 5\%$, $Z_{cri} = 1.96$
4. ***Test Statistic***

$$Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.5 - 0.541}{\sqrt{0.525 \times 0.475\left(\frac{1}{400} + \frac{1}{600}\right)}} = -1.269$$
$$|Z_{cal}| = 1.269$$

5. ***Conclusion:***

$$|Z_{cal}| < Z_{cri}$$

H_0*is accepted.*

3.In a city A, 20% of a random sample of 900 school boys has a certain slight physical defect. In another city B, 18.5% of a random sample of 1600 school boys has the same defect. Is the difference b/w the proportions significant at 5% L.O.S?

SOLUTION:

Given

$$n_1 = 900, n_2 = 1600$$
$$x_1 = 20\% \; of \; 900 = 180$$
$$x_2 = 18.5\% \; of \; 1600 = \frac{18.5}{100} \times 1600 = 296$$

$$P_1 = \frac{x_1}{n_1} = \frac{180}{900} = 0.2, P_2 = \frac{x_2}{n_2} = \frac{296}{1600} = 0.185$$

(i) ***Null hypothesis:*** $H_0: P_1 = P_2$
(ii) ***Alternative hypothesis:*** $H_1: P_1 \neq P_2$*(Two tailed test)*
(iii) ***L.O.S:*** $\alpha = 0.05(Z_{cri} = 1.96)$
(iv) ***Test statistic:***

$$|Z_{cal}| = \frac{|P_1 - P_2|}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$P = \frac{180 + 296}{900 + 1600} = 0.19$$

$$Q = 1 - P = 1 - 0.19 = 0.81$$

$$Z_{cal} = \frac{0.2 - 0.185}{\sqrt{(0.19)(0.81)\left(\frac{1}{900} + \frac{1}{1600}\right)}} = -0.918$$

5. ***Conclusion:***

$|Z_{cal}| < Z_{cri}$, *we accept* H_0.

Hence conclude that there is no significant difference b/w the proportions.

Confidence interval for the population proportion

1.Among 900 people in a state 90 are found to be Chapati eaters. Construct 99% confidence interval for the true proportion.

SOLUTION:

Given

$$n = 900, x = 90$$

What

$$p = \frac{x}{n} = \frac{90}{900} = 0.1$$

$$q = 1 - p = 1 - 0.1 = 0.9$$

The confidence interval / limits are

$$\left(p - Z_{\alpha/2}\sqrt{\frac{pq}{n}}, p + Z_{\alpha/2}\sqrt{\frac{pq}{n}}\right)$$

$$= \left(p - 3\sqrt{\frac{pq}{n}}, p + 3\sqrt{\frac{pq}{n}}\right)$$

$$= \left(0.1 - 3\sqrt{\frac{(0.1)(0.9)}{900}},\ 0.1 + 3\sqrt{\frac{(0.1)(0.9)}{900}}\right)$$
$$= [(0.1 - 3(0.01)),\ 0.1 + 3(0.01)]$$
$$= (0.07,\ 0.13)$$

Confidence Limits of mean of the population

$$\left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

2.A random sample of size 81 was taken whose variance is 20.25 & mean is 32. Construct 98% confidence interval.

SOLUTION:

Given

$$n = 81, \bar{x} = 32$$

Variance

$$\sigma^2 = 20.25 \Rightarrow \sigma = \sqrt{20.25} = 4.5$$

Also Given

$$Z_{\alpha/2} = 98\% = 0.02 = 2.33$$

What 98% Confidence interval is

$$\left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$
$$= \left[32 - 2.33\left(\frac{4.5}{\sqrt{81}}\right),\ 32 + 2.33\left(\frac{4.5}{\sqrt{81}}\right)\right]$$
$$= (30.83,\ 33.29)$$

3.The mean and standard deviation of a population are 11795 and 14054 respectively. Construct 95% confidence about the maximum error if $\bar{x} = 11795$and $n = 50$and also construct a 95% confidence interval for the true mean.

SOLUTION:

Given $\bar{x} = 11795,\ \sigma = 14054,\ n = 50$

95% confidence interval is,

$Z_{\alpha/2}$*at 95% Confidence interval is*

$$Z_{\alpha/2} = 1.96$$
$$\left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

$$= \left(11795 - 1.96\left(\frac{14054}{\sqrt{50}}\right),\ 11795 + 1.96\left(\frac{14054}{\sqrt{50}}\right)\right)$$
$$= (7899.4,\ 15690.6)$$

Max errorE = $Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ = $1.96\left(\frac{14054}{\sqrt{50}}\right)$ *=38.99*

4.A random sample size 16 values from a normal population stored a mean of 53 and a sum of squares of deviation from the mean equals to 150. Construct 95% confidence limits of mean of the population.

SOLUTION:

Given

$$n = 16$$
$$\bar{x} = 53,$$
$$\sigma^2 = 150 \Rightarrow \sigma = \sqrt{150} = 12.24$$
$$Z_{\alpha/2} = 95\% = 0.05 = 1.96$$

95% confidence interval is

$$\left(\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$
$$= \left(53 - (1.96)\frac{12.24}{\sqrt{16}},\ 53 + (1.96)\frac{12.24}{\sqrt{16}}\right)$$
$$= (47.003,\ 58.997)$$

Sampling Theory

Population:

Population is totality of statistical data forming a subject of investigation.

Ex:

1. *The population of heights of Indians.*
2. *The population of nationalised banks in India.*

Size of population:

Size of population is number of observations in population and is denoted by **"N"**.

Population may be classified into two types they are

(i) finite population.

(ii) infinite population.

(i) Finite Population:

The no. of observations in the population is finite then the population is called finite population.

(ii) Infinite Population:

The no. of observations in the populations is infinite then the population is called infinite population.

Ex:

The heights of residence in a certain city is a finite population.

Sampling:

The process of selection of a sample is called sampling.

Ex:

To assess the quality of a bag of rice, we examine only a portion of it by taking handful of it from the bag and then decide to purchase it (or) not.

Sample:

*A sample is a subset of population and the size of the sample is denoted by **"n"**.*

Samples are classified into two types. They are

1. *Small Samples*
2. *Large Samples*

1. Small Samples:

*If the no. of items in the sample (**n**) **< 30** items then the sample is called small sample.*

2. Large Samples:

*If the no. of items in the sample i.e **n ≥ 30** items then the sample is called large samples.*

Parameter and statistic:

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

Standard Error of a statistic:

*The standard error of statistic **'t'** is the standard deviation of the sampling distribution of the statistic. i.e., SE of sample mean is the SD of the sampling distribution of sample mean.*

Formulae for S.E:

→ *S.E of sample mean* $\bar{x} = \frac{\sigma}{\sqrt{n}}$

i.e. $S.E(\bar{x}) = \frac{\sigma}{\sqrt{n}}$

→ *S.E of sample proportion* $p = \sqrt{\frac{PQ}{n}}$

i.e. $S.E(p) = \sqrt{\frac{PQ}{n}}$

→ *S.E of the difference of two sample means* $\bar{x}_1 \& \bar{x}_2$

i.e.

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

→ *S.E of the difference of two proportions, i.e.*

$$S.E(P_1 - P_2) = \sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$$

Estimation:

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

Estimate:

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

Estimator:

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

Ex:

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

Types of Estimation:

→ ***Point Estimation:*** *If the estimate of the population parameter is given by a single value, then the estimate is called a point estimation of the parameter.*

→ ***Interval Estimation:*** *If the estimate of the population parameter is given by two different values b/w which the parameter may be considered to lie, then the estimate is called an interval estimation of the parameter.*

PROBLEM:

A *Population consists of 5 numbers* ***2, 3, 6, 8 and 11*** *consider all possible samples of size* ***2*** *which can be drawn (i) with replacement (ii) without replacement from this population. Find*

(a) *The mean of the population*

(b) *The standard Deviation of the population*

(c) *Mean of the Sampling distribution of Means*

(d) *The standard deviation of the sampling distribution of Mean*

SOLUTION:

Given population units are ***2, 3, 6, 8 and 11***

Population Size $N = 5$*and sample size* $n = 2$

(a) Mean of the population (μ ***)***

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

(b) Variance of the population (σ^2 ***)***

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

$$= \frac{1}{5}[(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2]$$
$$= \frac{1}{5}[16 + 9 + 0 + 4 + 25] = \frac{54}{5} = 10.8$$

Population S.D

$$\sigma = \sqrt{10.8} = 3.28$$

(c) Sampling with Replacement (Infinite population)

The total no. of sampling with replacement is

$$N^n = 5^2 = 25$$

25 samples of size = 2

Listing all possible samples of size 2 from population with replacement we get 25 samples

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

The sample means are

$$\frac{2+2}{2} = \frac{4}{2} = 2,\ \frac{2+3}{2} = \frac{5}{2} = 2.5,$$

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.

Mean of Sampling distribution of Means

$$\mu_{\bar{x}} = \frac{\textit{sum of all sample means}}{25}$$
$$= \frac{2 + 2.5 + 4 + 5 + \cdots + 7 + 9.5 + 11}{25}$$
$$= \frac{150}{25} = 6$$

Variance of Sampling Distribution of Mean

The variance of sampling distribution of mean is $\sigma_{\bar{x}}^2$

$$\sigma_{\bar{x}}^2 = \frac{1}{25}[(2-6)^2 + (2.5-6)^2 + \cdots] = 5.45$$
$$\sigma_{\bar{x}} = \sqrt{5.45} = 2.32$$

Verify the above result (Sampling with Replacement)

1. *Let*

$$\mu_{\bar{x}} = \mu = 6$$

2.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.29}{\sqrt{2}} = 2.32$$

Variance of Sampling Distribution of Mean

The variance of sampling distribution of mean is $\sigma_{\bar{x}}^2$

$$+(4.5-6)^2+(5.5-6)^2+(7-6)^2+(4-6)^2+(4.5-6)^2+(6-6)^2$$
$$+(7-6)^2+(8.5-6)^2+(5-6)^2+(5.5-6)^2+(7-6)^2+(8-6)^2$$
$$+(9.5-6)^2+(6.5-6)^2+(7-6)^2+(8.5-6)^2+(9.5-6)^2+(11-6)^2]$$
$$= \frac{1}{25}[136.25] = 5.45$$
$$\sigma_{\bar{x}} = \sqrt{5.45} = 2.32$$

Verify the above result (Sampling with Replacement)

1. *Let*

$$\mu_{\bar{x}} = \mu = 6$$

2.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.29}{\sqrt{2}} = 2.32$$

Sampling Without Replacement Process:

The total no. of samples sampling without Replacement is

$$^{N}C_n = \quad {}^{5}C_2 = 10$$

10 samples of size 2

Listing all possible samples of size 2 from population without replacement we get 10 samples

The samples are

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

The sample means are

2.5, 4, 5, 6.5, 4.5, 5.5, 7, 7, 8.5, 9.5

Mean of Sampling distribution of means

$$\mu_{\bar{x}} = \frac{\textit{Sum of all sample means}}{10}$$
$$= \frac{2.5+4+5+6.5+4.5+5.5+7+7+8.5+9.5}{10}$$
$$= \frac{60}{10} = 6$$

Variance of Sampling distribution of means is

$$\sigma_{\bar{x}}^2 = \frac{1}{10}[(2.5-6)^2 + (4-6)^2 + (5-6)^2 + (6.5-6)^2 + (4.5-6)^2 + (5.5-6)^2 + (7-6)^2 + (7-6)^2 + (8.5-6)^2 + (9.5-6)^2]$$

$$= \frac{40.50}{10} = 4.05$$

S.D

$$\sigma_{\bar{x}} = \sqrt{4.05} = 2.0125$$

Verify the result (Sampling without replacement)

(i)

$$\mu_{\bar{x}} = \mu = 6$$

(ii)

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}} = \sqrt{\frac{5-2}{5-1}\cdot\frac{10.8}{2}} = 2.0125$$

---

2. A random sample of size 100 is taken from an infinite population having the mean $\mu = 76$and variance $\sigma^2 = 256$. What is the probability that $\bar{x}$(sample mean) will be b/w 75 & 78.

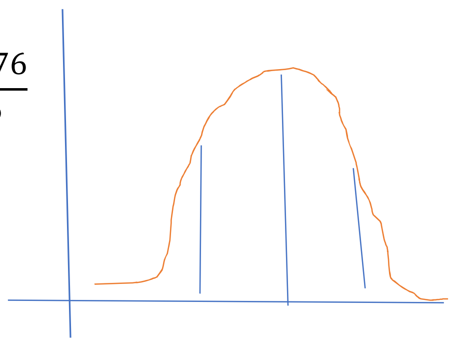
Given sample size $n = 100$, population mean $\mu = 76$

Variance $\sigma^2 = 256$

S.D $\sigma = \sqrt{256} = 16$

et

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{x}-76}{16/\sqrt{100}} = \frac{\bar{x}-76}{16/10} = \frac{\bar{x}-76}{1.6}$$



Now,

$$P(75 \le \bar{X} \le 78)$$

When **x̄ = 75**,

$$Z = \frac{75-76}{1.6} = -0.625$$

When **x̄ = 78**,

$$Z = \frac{78-76}{1.6} = 1.25$$

$$P(75 \le \bar{X} \le 78) = P(-0.625 \le Z \le 1.25)$$

$$= P(-0.625 \leq Z \leq 0) + P(0 \leq Z \leq 1.25)$$
$$= A(-0.625) + A(1.25)$$
$$= A(-0.63) + A(1.25)$$
$$= 0.2357 + 0.3944 = 0.6301$$

3) Find **95% confidence limits** for the mean of a normally distributed population from which the following sample was taken:

15, 17, 10, 18, 16, 9, 7, 11, 13, 14

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

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

$$= \frac{(15-13)^2 + (17-13)^2 + \cdots + (14-13)^2}{9} = \frac{40}{3}$$

$$s = \sqrt{\frac{40}{3}}$$

Since,

$$Z_{\alpha/2} = 1.96$$

$$Z_{\alpha/2}\frac{s}{\sqrt{n}} = 1.96\sqrt{\frac{40}{30}} = 2.26$$

Therefore, confidence limits are

$$\left(\bar{x} - Z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{s}{\sqrt{n}}\right)$$
$$= (13 - 2.26,\ 13 + 2.26)$$
$$= (10.74,\ 15.26)$$

4.A sample of size **400** is taken from a population whose **S.D. is 16**. Find the **standard error**.

Given:

Sample size **(n) = 400**

S.D. **(σ) = 16**

S.E of means

$$\text{S.E of means} = \frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{400}} = \frac{16}{20} = 0.8$$

4.How many different samples of size 2 can be chosen from a finite population of size 25.

Given $n = 2,\ N = 25$

Probability of choosing sample from the population

$$^{N}C_n = \ ^{25}C_2 = 300$$

5.Find the value of the finite population correction factor for $n = 10$, & $N = 100$

Given population size $N = 100$, sample size $n = 10$

Finite population correction factor

$$= \sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{100-10}{100-1}} = \sqrt{\frac{90}{99}}$$
$$= 0.9090$$

(6) A random sample of size 100 has a standard deviation of 5. Find maximum Error with 95% confidence.

Given $n = 100,\ \sigma = 5$

$$Z_{\alpha/2} \text{ at } 95\% \text{ Confidence} = 1.96$$

Maximum Error

$$= Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$
$$= 1.96 \times \frac{5}{\sqrt{100}} = 0.98$$

Practice Questions

Part A: One-Sample Mean (Z-Test)

1. *A random sample of **100** students has a mean height of **168 cm**. The population standard deviation is **8 cm**. Test whether the population mean height is **170 cm** at the **5% significance level**.*
2. *A factory claims that the average weight of a packet is **500 g**. A sample of **64** packets has a mean weight of **497 g**. The population standard deviation is **12 g**. Test the claim at the **1% significance level**.*
3. *A company claims that the average battery life is **50 hours**. A sample of **144** batteries has a mean life of **48.5 hours**. The population standard deviation is **6 hours**. Test the claim at the **5% level**.*

Part B: One-Sample Proportion (Z-Test)

4. *A manufacturer claims that **60%** of customers prefer its product. In a sample of **400** customers, **220** prefer the product. Test the claim at the **5% significance level**.*
5. *A university claims that **80%** of students pass an examination. In a sample of **500** students, **385** passed. Test the claim at the **1% significance level**.*

Part C: Difference Between Two Means

6. *Two machines produce bolts. From Machine A, a sample of **100** bolts has a mean length of **20.5 mm** with population standard deviation **2 mm**. From Machine B, a sample of **120** bolts has a mean length of **19.8 mm** with population standard deviation **2.5 mm**. Test whether the machines produce bolts of the same mean length at the **5% significance level**.*
7. *Two factories manufacture electric bulbs. A sample of **150** bulbs from Factory A has a mean life of **820 hours** (σ = **40 hours**). A sample of **200** bulbs from Factory B has a mean life of **810 hours** (σ = **36 hours**). Test whether there is a significant difference between the mean lifetimes at the **5% level**.*

Part D: Difference Between Two Proportions

8. *In a sample of **500** voters from City A, **280** support a candidate. In a sample of **600** voters from City B, **300** support the same candidate. Test whether the proportions are significantly different at the **5% level**.*
9. *Out of **400** students in College A, **280** passed an examination. Out of **500** students in College B, **325** passed. Test whether the pass percentages differ significantly at the **1% significance level**.*
10. *A survey shows that **180 out of 300** male voters and **240 out of 400** female voters support a policy. Test whether there is a significant difference between the proportions at the **5% significance level**.*

UNIT-5

Testing of Hypothesis – Small Sample

UNIT-5
Test of Hypothesis – Small Sample

Introduction

In statistics, Testing of Hypothesis for Small Samples is a method used to determine whether a claim about a population is true when the sample size is less than 30 (n < 30) and the population is normally distributed.
Since the sample size is small, the sample may not accurately represent the population. Therefore, special probability distributions such as the t-distribution, F-distribution, and Chi-square distribution are used instead of the normal (Z) distribution.
Small sample hypothesis testing helps researchers make reliable decisions even when only a limited amount of data is available.

Definition

Testing of Hypothesis for Small Samples is the statistical procedure used to test assumptions about population parameters when the sample size is less than 30 and the population variance is unknown.

Why Do We Use Small Sample Tests?

- When collecting large data is expensive.
- When experiments involve limited observations.
- When testing newly developed products.
- When medical trials involve a few patients.
- When research time is limited.
- To make decisions with limited available data.

Conditions for Small Sample Testing

- Sample size n < 30
- Population is approximately normally distributed
- Population variance is unknown
- Samples are selected randomly
- Observations are independent

Types of Small Sample Tests

Test	Purpose
Student's t-Test	Compare means
Paired t-Test	Compare before-after observations
Independent t-Test	Compare two independent sample means

Test	Purpose
F-Test	Compare two variances
Chi-square Test	Test population variance and goodness of fit

General Procedure of Hypothesis Testing

Step 1

State the Null Hypothesis (H_0)

Step 2

State the Alternative Hypothesis (H_1)

Step 3

Choose the Level of Significance (α)

Usually

- 5%
- 1%
- 10%

Step 4

Select appropriate Test Statistic

- t-test
- F-test
- χ^2-test

Step 5

Calculate Test Statistic

Step 6

Find Critical Value

Step 7

Decision

- Reject H_0
- Accept H_0 (Fail to Reject H_0)

Step 8

Write Conclusion

Advantages

✓ Useful when only limited data are available.
✓ Saves time and cost.
✓ Provides reliable decisions.
✓ Widely used in research.
✓ Suitable for laboratory experiments.
✓ Helps compare products, medicines and manufacturing processes.

Limitations

- Requires normal population.
- Sensitive to outliers.
- Less accurate if assumptions are violated.
- Small samples may not represent the entire population.

Real-Life Applications

1. Medical Research

A doctor wants to know whether a new medicine reduces blood pressure.
Only 12 patients are available for testing.
A paired t-test is used to compare blood pressure before and after treatment.

2. Manufacturing Industry

A company manufactures steel rods.
Engineers compare the variance of rod diameters produced by two different machines.
Since only 15 rods are tested from each machine,
F-test is used.

3. Education

A teacher introduces a new teaching method.
Marks of 20 students before and after teaching are compared.
A paired t-test determines whether the teaching method improved performance.

4. Agriculture

Scientists compare crop yield from two fertilizers.
Only 18 farms are available.
An independent t-test checks whether fertilizer A produces significantly higher yield than fertilizer B.

5. Pharmaceutical Industry

Researchers test whether a new tablet dissolves faster than the old tablet.
Only 10 samples are tested.
A t-test is applied.

6. Automobile Industry

Engineers compare fuel efficiency of two prototype engines.
Because only 8 prototype vehicles are available,
an independent t-test is performed.

7. Software Engineering

A company develops a new data compression algorithm.
Execution times of 15 programs before and after optimization are compared.
A paired t-test determines whether the optimization significantly reduces execution time.

---

8. Civil Engineering

Engineers compare the compressive strength of concrete prepared using two different cement brands.
Only 12 concrete cubes are tested.
An F-test checks whether the variability in strength differs, and a t-test compares the average strengths.

---

9. Electrical Engineering

A company manufactures LED bulbs.
Engineers test whether a new production process increases the average lifespan of bulbs.
Only 20 bulbs are tested.
A one-sample t-test is used.

---

10. Environmental Science

Researchers compare the pH level of river water before and after an industrial project.
Since only 14 water samples are collected,
a paired t-test is applied.

---

11. Artificial Intelligence (AI)

A machine learning engineer compares the accuracy of an old AI model and a new AI model using only 18 datasets.
A paired t-test determines whether the improvement is statistically significant.

---

12. Sports Analytics

A cricket coach evaluates whether a new training program improves batting performance.
Runs scored by 15 players before and after training are compared using a paired t-test.

Small sample: If the sample size $n \leq 30$, it is called a small sample.
The following are some important tests for small samples.

1. student–t-test
2. F-test
3. χ^2-test.

Degrees of freedom (D.O.F):

The number of independent variates which make up the statistic is known as the degrees of freedom (d.f) and it is denoted by ν.

Ex: In a set of data of $n$observations, if $k$is the no. of independent constants then

$$\nu = n - k.$$

Test of Hypothesis for Single Mean (μ):

Test statistic

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

where

$\bar{x}$= sample mean

σ= S.D

n= size

μ= population mean

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

if $s$is given directly.

Here Degree of freedom (D.f) $= n - 1$

Note:

1. For one tailed test | t_{cal} |= see t-table $(n{-}1)$ at α.
2. For two tailed tests | t_{cal} |= see t-table $(n{-}1)$ at $\alpha/2$.

Confidence interval for t:

Confidence interval for $t$is

$$\left(\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}},\ \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right)$$

Properties of t-distribution (or) Characteristics of t-distribution:

1. The shape of t-distribution is bell-shaped which is similar to that of a normal distribution and is symmetrical about the mean.
2. The t-distribution curve is also asymptotic to t-axis i.e., the two tails of the curve on both sides of $t = 0$extends to infinity.
3. It is symmetrical about the line $t = 0$.
4. The form of probability curve varies with degrees of freedom.
5. It is unimodal with Mean = Median = Mode.

6. The shape of t-distribution depends on sample size n.

Applications of t-distribution (or) uses of t-distribution:

1. To test the significance of the sample mean, when population variance is not given.
2. To test the significance of the mean of the sample i.e., to test if the sample mean differs significantly from the population mean.
3. To test the significance of the difference b/w two sample means.
4. To test the significance of an observed sample correlation coefficient and sample regression coefficient.

Student t-test for difference of means:

To test the significance difference b/w the sample means $\bar{X}$& $\bar{Y}$of 2 independent samples of sizes n_1& $n_2$with same variance, we use statistic

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$$s^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}$$

Here S_1& $S_2$are sample S.D.
t-distribution is followed with $n_1 + n_2 - 2$ d.o.f.

Problems:

1.The heights of 10 males of a given location are found to be 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches reasonable to believe that the average height is greater than 64 inches.

SOLUTION:

Sample
sample size $n = 10$
mean $\bar{x} = 66$
Population
mean $\mu = 64$inches
Let mean

$$\bar{x} = \frac{\Sigma X}{n} = \frac{70 + 67 + 62 + \cdots + 66}{10} = 66$$

Variance

$$\sigma^2 = \frac{1}{n-1}\Sigma(X_i - \bar{x})^2$$

$$= \frac{1}{10-1}[(70-66)^2 + (67-66)^2 + \cdots + (66-66)^2]$$
$$= \frac{1}{9}[90] = 10$$

S.D $(\sigma) = \sqrt{10} = 3.16$

(i) Null Hypothesis (H_0): $\mu = 64$

(ii) Alternative Hypothesis (H_1): $\mu > 64$(one tail)

(iii) LOS : $\alpha = 0.05$, df $= n - 1 = 10 - 1 = 9$

(iv) Test statistic :

$$t_{cal} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{66 - 64}{3.16/\sqrt{10}} = 1.9$$
$$t_{cri} = 1.833$$

(v)conclusion:

$$t_{cal} > t_{cri}$$

Reject H_0i.e $\mu > 64$

2.A sample of **26 bulbs** gives a mean life of **990 hrs** with a S.D of **20 hrs**. The manufacturer claims that the mean life of bulbs is **1000 hrs**. Identify the sample not upto the mark.

Solution:

$$H_0\text{: } \mu = 1000 \text{ hrs}$$
$$H_1\text{: } \mu < 1000 \text{ (one-tail)}$$
$$\alpha = 5\%,\ \alpha = 0.05$$

test statistic :

$$n = 26,\ \bar{x} = 990,\ \mu = 1000, \text{S.D}(s) = 20$$
$$t_{cal} = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{990 - 1000}{20/\sqrt{26}} = -2.5$$

(S.D of sample given directly)

$$|\ t_{cal}\ | = 2.5$$

Conclusion :

$t_{cri}$for $v = 26 - 1 = 25$d.f at α-level (one-tail)

$$t_{cri} = 1.708$$
$$t_{cal} > t_{cri}$$

Reject H_0, i.e $\mu < 1000$.

i.e the sample is not upto the mark.

2.A Random sample of **six steel beams**, has mean compressive strength of **58,392 P.S.i.** (pounds per square inch) with S.D of **648 P.S.i.** Use this information and the level of significance $\alpha = 0.05$to test whether the true average compressive strength of steel from which this sample came is **58,000 P.S.i.** Assume normality.

SOLUTION:

Given

sample size $n = 6$

Degrees of freedom $= n - 1 = 5$

sample $\bar{x} = 58{,}392$

(S.D) $= 648$

population mean $(\mu) = 58000$

1. Null Hypothesis H_0: $\mu = 58000$
2. A.H H_1: $\mu \neq 58000$(two tailed test)
3. LOS : $\alpha = 0.05$
4. Test statistic :

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

(i. S.D of sample given directly)

$$= \frac{58392 - 58000}{648/\sqrt{6}} = \frac{392}{648} \times \sqrt{5} = 1.353$$

$$|\ t_{cal}\ | = 1.353$$

5. Conclusion :

$$|\ t_{cri}\ | = 2.571$$

(i. See t-table $(n{-}1)$ at $\alpha/2$

" " 5 at $\frac{0.05}{2} = 0.025$)

$$|\ t_{cal}\ | < |\ t_{cri}\ |$$

∴ $H_0$is accepted

∴ The average compressive strength of steel beam is **58000 P.S.i**

3.Two horses **A** and **B** were tested according to the time to run a particular track with the following results.

Horse A	**28**	**30**	**32**	**33**	**33**	**29**	**34**
Horse B	29	30	30	24	27	29	

Test whether the two horses have the same running capacity.

Sol:

Given $n_1 = 7,\ n_2 = 6$

$$\bar{x} = \frac{\Sigma x}{n_1} = \frac{1}{7}(28+30+32+33+33+29+34) = \frac{219}{7} = 31.286$$

$$\bar{y} = \frac{\Sigma y}{n_2} = \frac{1}{6}(29+30+30+24+27+29) = \frac{169}{6} = 28.16$$

X	$x-\bar{x}$	$(x-\bar{x})^2$	**y**	$y-\bar{y}$	$(y-\bar{y})^2$
28	-3.286	10.8	29	0.84	0.7056
30	-1.286	1.6538	30	1.84	3.3856
32	0.714	0.51	30	1.84	3.3856
33	1.714	2.94	24	-4.16	17.3056
33	1.714	2.94	27	-1.16	1.3456
29	-2.286	5.226	29	0.84	0.7056
34	2.714	7.366			

$$\Sigma x = 219$$
$$\Sigma(x-\bar{x})^2 = 31.438$$
$$\Sigma y = 169$$
$$\Sigma(y-\bar{y})^2 = 26.8336$$

Now,

$$s^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2\right] = \frac{1}{11}(31.438+26.8336) = 5.23$$

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

(ii) Alternative Hyp. : H_1: $\mu_1 \neq \mu_2$

(iii) Test statistic :

$$t_{cal} = \frac{\bar{x}_1 - \bar{y}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{31.286-28.16}{2.3\sqrt{\frac{1}{7}+\frac{1}{6}}} = 2.443$$

$$\therefore\ t_{cal} = 2.443$$

(iv) Degrees of freedom $(v) = 7+6-2 = 11$

Tabulated value of $t$for 11 d.f at 5% LOS is **2.2**

(v) Conclusion :

$$|\, t_{cal}\, | > t_{cri}$$

we reject H_0.
Hence we conclude that both horses do not have the same running capacity.

2.Measurements of the fat content of two kinds of ice cream brand **A** and brand **B** yielded the following sample data.

Brand A	**13.5**	**14.0**	**13.6**	**12.9**	**13.0**
Brand B	12.9	13.0	12.4	13.5	12.7

Solution:
Step 1: State the hypotheses
Null Hypothesis

$$H_0: \mu_A = \mu_B$$

(There is no significant difference between the mean fat contents.)
Alternative Hypothesis

$$H_1: \mu_A \neq \mu_B$$

(There is a significant difference.)

Step 2: Calculate the sample means
For Brand A,

$$\bar{X}_1 = \frac{13.5 + 14.0 + 13.6 + 12.9 + 13.0}{5}$$
$$= \frac{67.0}{5} = 13.40$$

For Brand B,

$$\bar{X}_2 = \frac{12.9 + 13.0 + 12.4 + 13.5 + 12.7}{5}$$
$$= \frac{64.5}{5} = 12.90$$

Step 3: Calculate deviations and squares
Brand A

X	$X - \bar{X}$	$(X - \bar{X})^2$
13.5	0.1	0.01
14.0	0.6	0.36
13.6	0.2	0.04
12.9	-0.5	0.25

X	$X-\bar{X}$	$(X-\bar{X})^2$
13.0	-0.4	0.16

$$\sum(X-\bar{X})^2 = 0.82$$

Brand B

X	$X-\bar{X}$	$(X-\bar{X})^2$
12.9	0	0
13.0	0.1	0.01
12.4	-0.5	0.25
13.5	0.6	0.36
12.7	-0.2	0.04

$$\sum(X-\bar{X})^2 = 0.66$$

Step 4: Calculate pooled variance

$$S_p^2 = \frac{\sum(X-\bar{X})^2 + \sum(Y-\bar{Y})^2}{n_1+n_2-2}$$

$$= \frac{0.82+0.66}{5+5-2}$$

$$= \frac{1.48}{8} = 0.185$$

Hence,

$$S_p = \sqrt{0.185} = 0.430$$

Step 5: Calculate the t-statistic

Formula:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

Substitute the values:

$$= \frac{13.40-12.90}{0.430\sqrt{\frac{1}{5}+\frac{1}{5}}}$$

$$= \frac{0.50}{0.430\sqrt{0.4}}$$
$$= \frac{0.50}{0.430(0.6325)}$$
$$= \frac{0.50}{0.272}$$
$$t = 1.84$$

Step 6: Critical value

Degrees of freedom

$$n_1 + n_2 - 2 = 8$$

At **5% level (two-tailed)**,

$$t_{0.05,8} = 2.306$$

Step 7: Decision

Calculated value

$$t = 1.84$$

Table value

$$t = 2.306$$

Since

$$1.84 < 2.306$$

we **fail to reject** H_0.

Conclusion

There is **no significant difference** in the mean fat content of **Brand A** and **Brand B** at the **5% level of significance**.

Final Answer:

$\boxed{\text{The two ice cream brands do not differ significantly in their mean fat content.}}$

χ^2-Test for goodness of fit:

Suppose we are given a set of observed frequencies obtained under some experiment and we want to test if the experimental results support a particular hypothesis or theory. Karl pearson developed a test for testing the significance of discrepancy b/w experimental values and theoretical values obtained under some theory or hypothesis.

This test known as χ^2-test for of goodness of fit.

Karl pearson proved the test statistic.

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where

O = observed frequency (Experimental)

E = Expected frequency (Theoretical)

χ^2-Test is used to test whether the difference b/w observed and expected frequencies are significant.

Note:

1. If the data is given in series of n numbers then

$$\text{d.o.f} = n - 1$$

2. In case of Binomial distribution

$$\text{dof} = n - 1$$

Poisson

$$\text{dof} = n - 2$$

Normal

$$\text{dof} = n - 3$$

Properties of χ^2-distribution:

1. χ^2-distribution curve is not symmetrical, lies entirely in the first quadrant and hence not a normal curve since χ^2-values from 0to ∞.
2. Its depends on the degrees of freedom v.
3. If $v_1, v_2$are the degrees of freedom of $\chi_1^2$and χ_2^2then $(v_1+v_2)$is the degree of freedom of $\chi_1^2 + \chi_2^2$.
4. Mean = $v$and Variance = $2v$

Application of χ^2-distribution:

(i) To test the goodness of fit.

(ii) To test the independence of attributes.

(iii) To test if the population has a specified value of the variance σ^2.

PROBLEMS:

1.A pair of dice are thrown **360** times and frequency of each sum is indicated below.

Sum	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Frequency	8	24	35	37	44	65	51	42	26	14	14

Would you say that the dice are fair on the basis of the χ^2-test at **0.05 L.O.S.**

Sol:

1. Null Hypothesis (H_0): The dice are fair.
2. Alternative Hypothesis (H_1): The dice are not fair.
3. L.O.S :

$$\alpha = 5\% = 0.05$$

The probabilities of a getting sum 2,3,4,5,6,7,8,9,10,11,12are

x	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
$P(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$Ef(x)$	10	20	30	40	50	60	50	40	30	20	10

Observed frequency (O)

Expected frequency

$$E = 360 \times P(x)$$

Sum	**Observed frequency (O)**	**Expected frequency (E)**	$(O-E)^2$	$\frac{(O-E)^2}{E}$
2	8	10	4	0.4
3	24	20	16	0.8
4	35	30	25	0.833
5	37	40	9	0.225
6	44	50	36	0.72
7	65	60	25	0.417
8	51	50	1	0.02
9	42	40	4	0.1
10	26	30	16	0.53

Sum	Observed frequency (O)	Expected frequency (E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
11	14	20	36	1.8
12	14	10	16	1.6

$$\sum\frac{(O-E)^2}{E} = 7.445$$

$$\chi^2_{cal} = \sum\frac{(O-E)^2}{E} = 7.445$$

Conclusion:

Degrees of freedom $= n-1 = 11-1 = 10$

$$|\ \chi^2_{cri}\ |= 18.3$$

$$|\ \chi^2_{cal}\ |<|\ \chi^2_{cri}\ |$$

Accept H_0

i.e The dice are fair

($\because$ see χ^2-table 10 d.f at 0.05 = 18.307)

---

2. Fit a poisson distribution to the following data and test the goodness of fit at 0.05 LOS.

x	0	1	2	3	4	5	6	7
$f(x)$	305	366	210	80	28	9	2	1

Sol:

1. Null Hypothesis H_0: Poisson distribution is a good fit
2. H_1: Poisson distribution is not a good fit
3. L.O.S : $\alpha = 0.05$

Mean (λ)

$$= \frac{\sum xf(x)}{\sum f(x)} = \frac{0\times305+1\times366+2\times210+3\times80+4\times28+5\times9+6\times2+7\times1}{1001}$$

$$= \frac{1202}{1001} = 1.2$$

By poisson distribution

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0,1,2,\dots,7$$

$$= \frac{e^{-1.2}(1.2)^x}{x!}$$

Expected frequencies

$$E = N \times P(x) = 1001 \times \frac{e^{-1.2}(1.2)^x}{x!}$$
$$= 1001 \times 0.3\frac{(1.2)^x}{x!} = \frac{300.3(1.2)^x}{x!}$$

x	Observed frequency (O)	Expected frequency (E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
0	305	300.3	22.09	0.0736
1	366	360.36	31.810	0.0883
2	210	216.216	38.639	0.1787
3	80	86.48	41.990	0.4357
4	28	25.946	4.249	0.1626
5	9	6.227	7.684	1.2348
6	2	1.245	0.570	0.4578
7	1	0.213	0.619	2.9061

$$\sum f = 1001$$
$$\chi^2 = \sum \frac{(O-E)^2}{E} = 5.5874$$

Conclusion:

D.O.F $= n - 2 = 8 - 2 = 6$

Tabulated value of χ^2at 5% level of 6 d.f = **12.592**

$$|\chi^2_{cal}| < |\chi^2_{cri}|$$

$H_0$is accepted.

i.e Poisson distribution is a good fit.

χ²-test for Independent of Attributes:

Attribute:

An Attribute means a quality or characteristic. Examples of attribute are climbing, smiling, honesty, beauty etc.

An attribute may be marked by its presence (position) or absence in a no. of a given population.

Let us consider two attributes A and B.

A is divided into two classes and B is divided into two classes. The various cell frequencies can be expressed in the following table known as **2×2 contingency table.**

A	A	b
B	C	d
	a+c	b+d

A	a	b	a+b
B	c	d	C+d

The Expected frequencies are given by

$E(a) = \frac{(a+c)(a+b)}{N}$	$E(b) = \frac{(b+d)(a+b)}{N}$	a+b
$(c) = \frac{(a+c)(c+d)}{N}$	$E(d) = \frac{(b+d)(c+d)}{N}$	c+d
a+c	b+d	$N = a+b+c+d$

Note:

In this χ^2-test, The test if two attributes A & B under consideration are independent or not.

Null Hypothesis H_0: Attributes are independent.

$$\text{D.O.F} = (r-1)(s-1)$$

where

r= no. of rows

s= no. of columns

---

1)

On the basis of information given below about the treatment of **200 patients** suffering from a disease, state whether the new treatment is comparatively superior to the conventional treatment.

	favourable	**Not favourable**	**Total**
New	60	30	90
Conventional	40	70	110

Solution:

(i) Null Hypothesis H_0: No difference b/w new and conventional treatment (or) New and Conventional treatment are independent.

(ii) Alternative Hypothesis (H_1): New and conventional treatments are **not** independent.

(iii) L.O.S :

$$\alpha = 5\% = 0.05$$

Test statistic :

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

(iv) Computation :

	Favourable	**Not favourable**	**Total**
New	60	30	90
Conventional	40	70	110
Total	100	100	200

Table of Expected frequencies

	Favourable	**Not favourable**	**Total**
New	$\frac{100 \times 90}{200} = 45$	$\frac{100 \times 90}{200} = 45$	90
Conventional	$\frac{100 \times 110}{200} = 55$	$\frac{100 \times 110}{200} = 55$	110
Total	100	100	200

observed frequency (O)	**Expected frequency (E)**	$(O-E)^2$	$\frac{(O-E)^2}{E}$
60	45	225	5
30	45	225	5
40	55	225	4.09
70	55	225	4.09

$$\sum \frac{(O-E)^2}{E} = 18.18$$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 18.18$$

$$|\chi^2_{cal}| = 18.18$$

Conclusion:

D.O.F

$$= (r-1)(s-1) = (2-1)(2-1) = 1$$
$$|\,\chi^2_{cri}\,| = 3.8411$$

($\because$ see χ^2-table 1 at $0.05 = 3.8411$)

$$|\,\chi^2_{cal}\,| > |\,\chi^2_{cri}\,|$$

$\therefore H_0$ is rejected.

i.e New and conventional treatment are not independent.

2.Given the following contingency table for hair colour & eye colour. Find the value of χ^2. Is there good association b/w the two?

Eye Colour	Fair	Brown	Black	Total
Blue	15	5	20	40
Grey	20	10	20	50
Brown	25	15	20	60
Total	60	30	60	150

Sol:

write Given table.

(i) Null Hypothesis : H_0: The two attributes hair & eye colour are independent.

(ii) H_1: The 2 attributes are not independent.

(iii) LOS :

$$\alpha = 0.05 \ (5/100)$$
$$\text{d.o.f} = (r-1)(s-1)$$

where r =no. of rows

s =no. of columns

$$= (3-1)(3-1)$$
$$= (2)(2) = 4$$

(iv) Test statistic:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Computations:

write given table.

Table of Expected frequencies:

Eye Colour	Fair	Brown	Black	Total
Blue	$\frac{60\times40}{150}=16$	$\frac{30\times40}{150}=8$	$\frac{60\times40}{150}=16$	**40**
Grey	$\frac{60\times50}{150}=20$	$\frac{30\times50}{150}=10$	$\frac{60\times50}{150}=20$	**50**
Brown	$\frac{60\times60}{150}=24$	$\frac{30\times60}{150}=12$	$\frac{60\times60}{150}=24$	**60**
Total	**60**	**30**	**60**	**150**

Observed frequency (O)

O	Expected frequency (E)	(O–E)	(O–E)²	(O–E)² / E
15	16	-1	1	0.0625
5	8	-3	9	1.125
20	16	4	16	1
20	20	0	0	0
10	10	0	0	0
20	20	0	0	0
25	24	1	1	0.042
15	12	3	9	0.75
20	24	-4	16	0.666

$$\chi^2_{cal} = \sum \frac{(O-E)^2}{E} = 3.6485$$

Conclusion:

χ^2_{cri} value for d.o.f $=(3-1)(3-1)=4$ at 5% L.O.S is 9.488

$$\chi^2_{cal} < \chi^2_{cri}$$

so we accept H_0.

i.e The hair colour & eye colour are independent.

From the following data find whether there is any significant liking in the habit of taking soft drinks among the categories of employees.

Soft drinks	Clerks	Teachers	Officers
Pepsi	10	25	65
Thumsup	15	30	65
Fanta	50	60	80

Solution:

tep 1: Null and Alternative Hypotheses

Null Hypothesis (H_0):

There is **no significant association** between the type of employee and the preference for soft drinks.

Alternative Hypothesis (H_1):

There **is a significant association** between the type of employee and the preference for soft drinks.

Step 2: Level of Significance

$$\alpha = 0.05$$

Step 3: Observed Frequencies (O)

Soft Drinks	Clerks	Teachers	Officers	Row Total
Pepsi	10	25	65	100
Thumsup	15	30	65	110
Fanta	50	60	80	190
Column Total	**75**	**115**	**210**	**400**

Grand Total

$$N = 400$$

Step 4: Expected Frequencies

Formula

$$E = \frac{(\text{Row Total})(\text{Column Total})}{\text{Grand Total}}$$

Pepsi

$$E_{11} = \frac{100 \times 75}{400} = 18.75$$
$$E_{12} = \frac{100 \times 115}{400} = 28.75$$
$$E_{13} = \frac{100 \times 210}{400} = 52.5$$

Thumsup

$$E_{21} = \frac{110 \times 75}{400} = 20.625$$
$$E_{22} = \frac{110 \times 115}{400} = 31.625$$
$$E_{23} = \frac{110 \times 210}{400} = 57.75$$

Fanta

$$E_{31} = \frac{190 \times 75}{400} = 35.625$$
$$E_{32} = \frac{190 \times 115}{400} = 54.625$$
$$E_{33} = \frac{190 \times 210}{400} = 99.75$$

Step 5: Calculate $(O-E)^2/E$

Soft Drink	O	E	O–E	(O–E)²/E
Pepsi-Clerks	10	18.75	-8.75	4.083
Pepsi-Teachers	25	28.75	-3.75	0.489
Pepsi-Officers	65	52.50	12.50	2.976
Thumsup-Clerks	15	20.625	-5.625	1.534
Thumsup-Teachers	30	31.625	-1.625	0.084
Thumsup-Officers	65	57.75	7.25	0.911
Fanta-Clerks	50	35.625	14.375	5.801
Fanta-Teachers	60	54.625	5.375	0.529
Fanta-Officers	80	99.75	-19.75	3.911

Sum of Chi-square values

$$\chi^2_{cal} = 4.083 + 0.489 + 2.976 + 1.534 + 0.084 + 0.911 + 5.801 + 0.529 + 3.911$$
$$\chi^2_{cal} = 20.318$$

Step 6: Degrees of Freedom

$$(r-1)(c-1)$$

where

$$r = 3, c = 3$$
$$= (3-1)(3-1)$$
$$= 2 \times 2 = 4$$

Step 7: Critical Value

From Chi-square table,

At **5% level of significance**

Degrees of freedom = 4

$$\chi^2_{0.05,4} = 9.488$$

Step 8: Decision

Since

$$\chi^2_{cal} = 20.318$$

and

$$\chi^2_{tab} = 9.488$$
$$\chi^2_{cal} > \chi^2_{tab}$$

Therefore,

Reject H_0.

Conclusion

There **is a significant liking in the habit of taking soft drinks among the categories of employees.**

Hence, the preference for **Pepsi, Thumsup, and Fanta depends on the category of employees (Clerks, Teachers, and Officers).**

Final Answer:

$$\boxed{\begin{array}{c}\chi^2_{cal} = 20.318,\ \chi^2_{tab} = 9.488,\\ \text{Reject } H_0\end{array}}$$

Therefore, there is a significant association between employee category and soft drink preference.

.

χ²-test for population Variance:

Suppose that a random sample is drawn from a normal population with mean μ and variance σ², under the null hypothesis H_0that the population variance σ² has a specified value σ_0^2.

Test statistic is

$$\chi^2 = \frac{\sum(x_i - \bar{x})^2}{\sigma_0^2} = \frac{ns^2}{\sigma_0^2}$$

It follows χ²-distribution with $(n-1)$d.o.f.

1.

A firm manufacturing rivets wants to limit variations in their length as much as possible. The lengths (in cms) of rivets manufactured by new process are

2.15 1.99 2.05 2.12 2.17

2.01 1.98 2.03 2.25 1.93

Examine whether the new process can be considered superior to the old if the old population has S.D 0.148 cm?

Solution:

Given $n = 10,\ \bar{x} = \frac{\sum x_i}{n} = \frac{20.68}{10} = 2.068$

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{0.09096}{9} = 0.0101$$

and $\sigma_0 = 0.148$

(i) Null Hypothesis H_0: $\sigma^2 = \sigma_0^2$

$$H_1:\ \sigma^2 \neq \sigma_0^2$$

$$\alpha = 0.05$$

$$\chi^2 = \frac{ns^2}{\sigma_0^2} = \frac{10 \times 0.0101}{(0.148)^2} = 4.8$$

Conclusion:

d.o.f $= n - 1 = 9$

$$|\ \chi^2_{cri}\ | = 16.919$$

$$|\ \chi^2_{cal}\ | < |\ \chi^2_{cri}\ |$$

$H_0$is accepted.

2.

A random sample of size 20 from a normal population gave a mean of 42 and variance 25. Test the hypothesis that the population S.D is 8 at 5% level.

$$n = 20,\ \bar{x} = 42,\ s^2 = 25,\ \sigma_0 = 8$$

$$\chi^2 = \frac{20 \times 25}{8^2} = 7.8125, \chi^2_{cri} = 30.14$$

V=n-1=20-1=19

$$\chi^2_{cri} = 8.907$$

$$|\ \chi^2_{cal}\ |>|\ \chi^2_{cri}\ |$$

$H_0$is rejected

F-TEST:

This is very useful in testing the equality of population means by comparing sample variances.

To test whether there is any significant difference b/w two estimates of population variance (or) To test if the 2 samples have came from the same population, we use F-test. In this case we setup null hypothesis,

$$H_0: \sigma_1^2 = \sigma_2^2$$

i.e., population variances are same.

The test statistic is

$$F = \frac{S_1^2}{S_2^2}$$

where

$$S_1^2 = \frac{\sum(x_i - \bar{x})^2}{n_1 - 1}, S_2^2 = \frac{\sum(y_i - \bar{y})^2}{n_2 - 1}$$

and

$$S_1^2 > S_2^2$$

The d.o.f are

$$V_1 = n_1 - 1, V_2 = n_2 - 1$$

n_1= first sample size, n_2= second sample size.

Properties of F-distribution:

1. F-distribution curve lies entirely in 1st quadrant.

2. The F-curve depends not only on 2 parameters v_1& $v_2$but also on the order in which they are stated.
3. If sample variance $S^2$is given, we can obtain population variance $\sigma^2$by using the relation,

$$n\sigma^2 = (n-1)S^2$$

& vice versa.

4. $F_\alpha(v_1, v_2)$is the value of F with v_1& v_2d.o.f such that the area under F-distribution curve to the right of F is α.
5. The mode of F-distribution is less than unity.
6. F-distribution (or) F-test is not symmetric.

The following random samples are measurements of the heat producing capacity (in millions of calories per ton) of specimen of coal from two mines.

Mine-I	**8260**	**8730**	**8830**	**8070**	**8340**	**-**
Mine-II	7950	7890	7900	8140	7220	7840

Use **0.01 level of significance** test whether variances of two samples are equal?

SOLUTION:

1. Null Hypothesis:

We state the null Hypothesis as H_0

$$H_0\text{: } \sigma_1^2 = \sigma_2^2$$

(or)

$$\sigma_1^2 = \sigma_2^2$$

2. Alternative Hypothesis:

$$H_1\text{: } \sigma_1^2 \neq \sigma_2^2$$

(or)

$$\sigma_1^2 \neq \sigma_2^2$$

3. Level of Significance (L.O.S):

$$\alpha = 0.01$$

4. Degrees of freedom (d.o.f):

$$V_1 = n_1 - 1 = 5 - 1 = 4$$
$$V_2 = n_2 - 1 = 6 - 1 = 5$$

5. Test statistic:

$$F = \frac{S_1^2}{S_2^2}$$

$$S_1^2 = \frac{\sum(x_i - \bar{x})^2}{n_1 - 1}, S_2^2 = \frac{\sum(y_i - \bar{y})^2}{n_2 - 1}$$

$$\bar{x} = \frac{8260 + 8130 + 8350 + 8070 + 8340}{5} = 8230$$

$$\bar{y} = \frac{7950 + 7890 + 7900 + 8140 + 7920 + 7840}{6} = 7940$$

Null Hypothesis:-

We state the null Hypothesis as H_0

$H_0 : \mu_1 = \mu_2$ (or) $\sigma_1^2 = \sigma_2^2$

Alternative Hypothesis:-

$H_1 : \mu_1 \neq \mu_2$ (or) $\sigma_1^2 \neq \sigma_2^2$

3. Level of Significance (L.O.S):-

$\alpha = 0.01$

4. Degrees of freedom (d.o.f):-

$V_1 = n_1 - 1 = 5 - 1 = 4$

$V_2 = n_2 - 1 = 6 - 1 = 5$

5. Test Statistic:-

$$F = \frac{S_1^2}{S_2^2}$$

$$S_1^2 = \frac{\sum(x_i - \bar{x})^2}{n_1 - 1}, S_2^2 = \frac{\sum(y_i - \bar{y})^2}{n_2 - 1}$$

$$\bar{x} = \frac{8260 + 8130 + 8350 + 8070 + 8340}{4} = 8230$$

$$\bar{y} = \frac{7950 + 7890 + 7900 + 8140 + 7920 + 7840}{5} = 7940$$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
8260	30	900	7950	10	100
8130	100	10000	7890	50	2500
8350	120	14400	7900	40	1600
8070	160	25600	8140	200	40000
8340	110	12100	7920	20	400
			7840	100	10000
		$\Sigma(x_i - \bar{x})^2 = 63900$			**$\Sigma(y_i - \bar{y})^2 = 54600$**

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$

$$S_1^2 = \frac{63900}{4} = 15750$$
$$S_2^2 = \frac{54600}{5} = 10920$$
$$F = \frac{S_1^2}{S_2^2} = \frac{15750}{10920} = 1.44$$

Fcal = 1.44

Conclusion:-

For at 1% LOS, for (4,5) d.o.f

$$F_{cv} = 11.39$$

Fcal < Fcri ∴ accept H_0.

Hence we conclude that $\boldsymbol{\sigma_1^2 = \sigma_2^2}$

The measurements of the output of two units have given the following results. Assuming the both samples have been obtained from the populations at 10% significant level, test whether two populations have the same variance.

Unit-A	**14.1**	**10.1**	**14.7**	**13.7**	**14.0**
Unit-B	14.0	14.5	13.7	12.7	14.1

Sol:

Given $n_1 = 5$, $n_2 = 5$

$$\bar{x} = \frac{\sum x_i}{n_1} = \frac{1}{5}(14.1 + 10.1 + 14.7 + 13.7 + 14) = 13.32$$
$$\bar{y} = \frac{\sum y_i}{n_2} = \frac{1}{5}(14 + 14.5 + 13.7 + 12.7 + 14.1) = 13.8$$

(i) Null Hypothesis:

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

$$\alpha = 10\% = 0.01$$

Test statistic

$$F = \frac{S_1^2}{S_2^2}(S_1 > S_2) = \frac{3.372}{0.46} = 7.33$$

Complete Calculation for mean & S.D's of samples:

x_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	y_i	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$
14.1	0.78	0.6084	14.0	0.2	0.04
10.1	-3.22	10.3684	14.5	0.7	0.49
14.7	1.38	1.9044	13.7	-0.1	0.01
13.7	0.38	0.1444	12.7	-1.1	1.21
14.0	0.68	0.4624	14.1	0.3	0.09

$$\sum(x_i-\bar{x})^2 = 13.488$$

$$S_1^2 = \frac{\sum(x_i-\bar{x})^2}{n_1-1} = \frac{13.488}{4} = 3.372$$

$$S_2^2 = \frac{\sum(y_i-\bar{y})^2}{n_2-1} = \frac{1.84}{4} = 0.46$$

v. Conclusion :

D.of freedom $(n_1-1, n_2-1) = (4,4)$

Tabulated value for **F** at **10% L.O.S** for $(4,4)$d.o.f is **15.98**

$$F_{Cri} = 15.98$$

$$|\, F_{Cal}\, | < |\, F_{Cri}\, |$$

$\therefore H_0$is accepted.

i.e There is no significant difference b/w the variances.

Two independent samples of sizes 9 and 7 from a normal population had the following values of the variables

Sample I	18	13	12	15	12	14	16	14	15
Sample II	16	19	13	16	18	13	15		

Do the estimates of the population variance differ significantly at 5% level?

SOLUTION:

Given: $n_1 = 9,\ n_2 = 7$

1. **Null Hypothesis** H_0: $\sigma_1^2 = \sigma_2^2$
2. **A.H** (H_1) **:** $\sigma_1^2 \neq \sigma_2^2$
3. **L.O.S** (α) **:** $\alpha = 0.05$, d.o.f $v_1 = n_1 - 1 = 9 - 1 = 8,\ v_2 = n_2 - 1 = 7 - 1 = 6$
4. **Test statistic :**

$$F = \frac{S_1^2}{S_2^2}$$

$$\bar{x} = \frac{18+13+12+15+12+14+16+14+15}{9} = \frac{129}{9} = 14.33$$

$$\bar{y} = \frac{16+19+13+16+18+13+15}{7} = \frac{110}{7} = 15.71$$

$$S_1^2 = \frac{\sum(x_i - \bar{x})^2}{n_1 - 1}$$

$$= \frac{(18-14.33)^2+(13-14.33)^2+(12-14.33)^2+(15-14.33)^2+(12-14.33)^2+(14-14.33)^2+(16-14.33)^2+(14-14.33)^2+(15-14.33)^2}{9-1}$$

$$= \frac{13.47+1.77+5.43+0.45+5.43+0.11+2.79+0.11+0.45}{8}$$

$$= \frac{29.99}{8} = 3.75$$

$$= \frac{(16-15.71)^2+(19-15.71)^2+(13-15.71)^2+(16-15.71)^2+(18-15.71)^2+(13-15.71)^2+(15-15.71)^2}{7-1}$$

$$= \frac{31.40}{6} = 5.23$$

$$F_{Cal} = \frac{S_1^2}{S_2^2} = \frac{3.75}{5.23} = 0.72$$

Conclusion :

D.O.F : $V_1 = n_1 - 1 = 9 - 1 = 8,\ \ V_2 = n_2 - 1 = 7 - 1 = 6$

$$F_{Cri} = 4.15$$

(C: See F table d.o.f $(8, 6)$*at 5% level)*

$$|\ F_{Cal}\ | < |\ F_{Cri}\ |$$

Accept H_0.

i.e There is no significant difference b/w the variances.

A group of 10 rats fed on diet A and another group of 8 rats fed on diet B recorded the following increase in weight

Diet A	*5*	*6*	*8*	*1*	*12*	*4*	*3*	*9*	*6*	*10*
Diet B	*2*	*3*	*6*	*8*	*10*	*1*	*2*	*8*		

Find if the variance are significantly different.

1. ***N.H*** (H_0) ***:*** $\sigma_1^2 = \sigma_2^2$
2. ***A.H*** (H_1) ***:*** $\sigma_1^2 \neq \sigma_2^2$
3. ***L.O.S :*** $\alpha = 0.05$
4. ***Test Statistic :***

$$F = \frac{S_1^2}{S_2^2}, n_1 = 10,\ n_2 = 8$$

$$\bar{x} = \frac{5+6+8+1+12+4+3+9+6+10}{10} = \frac{64}{10} = 6.4,$$

$$\bar{y} = \frac{2+3+6+8+10+1+2+8}{8} = \frac{40}{8} = 5$$

$$S_1^2 = \frac{\sum (x_i - \bar{x})^2}{n_1 - 1}$$

$$= \frac{(5-6.4)^2 + (6-6.4)^2 + (8-6.4)^2 + (1-6.4)^2 + (12-6.4)^2 + (4-6.4)^2 + (3-6.4)^2 + (9-6.4)^2 + (6-6.4)^2 + (10-6.4)^2}{10-1}$$

$$= \frac{1.96 + 0.16 + 2.56 + 29.16 + 31.36 + 5.76 + 11.56 + 6.76 + 0.16 + 12.96}{9}$$

$$= \frac{102.40}{9} = 11.38$$

$$S_2^2 = \frac{\sum (y_i - \bar{y})^2}{n_2 - 1}$$

$$= \frac{(2-5)^2 + (3-5)^2 + (6-5)^2 + (8-5)^2 + (10-5)^2 + (1-5)^2 + (2-5)^2 + (8-5)^2}{7}$$

$$= \frac{9+4+1+9+25+16+9+9}{7} = \frac{82}{7} = 11.71$$

$$F_{Cal} = \frac{S_1^2}{S_2^2} = \frac{11.38}{11.71} = 0.97$$

Conclusion :

$$V_1 = 10 - 1 = 9, V_2 = 8 - 1 = 7$$
$$F_{Cri} = 3.68$$

(.: see F table at $(V_1, V_2) = (9,7)$ *d.o.f at* $\alpha = 0.05$*)*

$$|\ F_{Cal}\ | < |\ F_{Cri}\ |$$

Accept H_0.

i.e There is no significance difference b/w the variances.

Practice Questions – Small Sample Tests

Part A – One-Sample t-Test

Q1. *A random sample of 8 observations is:*
18, 20, 17, 19, 22, 21, 20, 18.
*Test whether the population mean is **19** at the **5% level of significance**.*

Q2. *A sample of 10 students obtained the following marks:*
62, 65, 68, 70, 66, 69, 64, 67, 71, 63.
*Test whether the average mark is **65** at the **5% significance level**.*

Q3. *A manufacturer claims that the average life of bulbs is* ***1000 hours****. A sample of 9 bulbs gave:*
980, 1020, 995, 1015, 990, 1005, 1010, 985, 1000.
Test the claim at the ***5% level****.*

Part B – Two-Sample Independent t-Test

Q4. *Two independent samples are:*
Sample A: 45, 48, 50, 47, 49
Sample B: 42, 44, 46, 45, 43
Test whether the population means are equal at the ***5% significance level****.*

Q5. *The weights (kg) of two groups of students are:*
Group I: 58, 60, 62, 59, 61
Group II: 55, 57, 56, 58, 54
Test whether there is a significant difference between the mean weights.

Q6. *Machine A produces:*
20, 22, 21, 23, 24
Machine B produces:
18, 19, 20, 21, 22
Test whether both machines have the same average output.

Part C – Paired t-Test

Q7. *The following are marks before and after coaching.*

Student	Before	After
1	45	50
2	50	54
3	48	52
4	52	55
5	47	51
6	49	53

Test whether coaching improved performance.

Q8. *Blood pressure readings before and after treatment are:*
Before: 150, 145, 160, 155, 148
After: 142, 140, 150, 148, 144
Test whether the treatment is effective.

Part D – F-Test (Equality of Variances)

Q9. *Two samples have:*
Sample I: $n = 8$, $s^2 = 36$

Sample II: $n = 10$, $s^2 = 16$

Test whether the population variances are equal at the ***5% significance level****.*

Q10. *Two machines produce the following variances:*

Machine A: $n = 12$, $s^2 = 25$

Machine B: $n = 10$, $s^2 = 16$

Test whether there is any significant difference in their variances.

Part E – Chi-Square Test for Goodness of Fit

Q1. *A die was thrown 120 times with the following results:*

Face	***1***	***2***	***3***	***4***	***5***	***6***
Frequency	*15*	*20*	*18*	*22*	*25*	*20*

Test whether the die is fair at the ***5% level of significance****.*

Part F – Chi-Square Test of Independence

Q4. *The following table gives the smoking habits of employees.*

Smoking Habit	***Male***	***Female***
Smoker	*60*	*30*
Non-Smoker	*40*	*70*

Test whether smoking habit is independent of gender.

Q10. *A survey of 200 students gave the following information:*

Transport	***Boys***	***Girls***
Bus	*50*	*40*
Bicycle	*30*	*20*
Walk	*20*	*40*

Test whether the mode of transport is independent of gender.

★ Final Inspirational Quote:

"Statistics uncovers the truth hidden in data, while Numerical Methods transform complex problems into practical solutions."

THE END