QUANTITATIVE METHODS FOR ECONOMICS II

COURSE CODE: M23EC10DC

Postgraduate Programme in Economics

Discipline Core Course

Self Learning Material

SREENARAYANAGURU OPEN UNIVERSITY

The State University for Education, Training and Research in Blended Format, Kerala

SREENARAYANAGURU OPEN UNIVERSITY

Vision

To increase access of potential learners of all categories to higher education, research and training, and ensure equity through delivery of high quality processes and outcomes fostering inclusive educational empowerment for social advancement.

Mission

To be benchmarked as a model for conservation and dissemination of knowledge and skill on blended and virtual mode in education, training and research for normal, continuing, and adult learners.

Pathway

Access and Quality define Equity.

Quantitative Methods for Economics II

Course Code: M23EC10DC

Semester - III

Discipline Core Course
Postgraduate Programme in Economics
Self Learning Material

SREENARAYANAGURU
OPEN UNIVERSITY

SREENARAYANAGURU OPEN UNIVERSITY

The State University for Education, Training and Research in Blended Format, Kerala

വിദ്യകൊണ്ട് സ്വതന്ത്രരാവുക

SREENARAYANAGURU
OPEN UNIVERSITY

QUANTITATIVE METHODS FOR ECONOMICS II

Course Code: M23EC10DC
Semester- III
Discipline Core Course
Postgraduate Programme in Economics

Academic Committee

Dr. Anitha V
Santhosh T Varghese
Dr. Prasad A.K.
Dr. B. Pradeepkumar
Dr. C.C. Babu
Dr. Sindhu Prathap
Dr. Christabella P. J.
Dr. Aparna Das
Dr. Moti George
Dr. S. Jayasree

Development of the Content

Dr Anitha C.S.
Dr. Sanoop M.S.

Review and Edit

Dr. G. Hari Prakash

Linguistics

Dr. Anitha C.S.

Scrutiny

Yedu T. Dharan
Dr. Suchithra K.R.
Soumya V.D.
Muneer K.
Dr. Smitha K.

Design Control

Azeem Babu T.A.

Cover Design

Jobin J.

Co-ordination

Director, MDDC :
Dr. I.G. Shibi
Asst. Director, MDDC :
Dr. Sajeevkumar G.
Coordinator, Development:
Dr. Anfal M.
Coordinator, Distribution:
Dr. Sanitha K.K.

Scan this QR Code for reading the SLM on a digital device.

Edition
April 2025

ISBN 978-81-985949-4-5

9 788198 594945

www.sgou.ac.m

YouTube

Visit and Subscribe our Social Media Platforms

MESSAGE FROM VICE CHANCELLOR

Dear learner,

I extend my heartfelt greetings and profound enthusiasm as I warmly welcome you to Sreenarayanaguru Open University. Established in September 2020 as a state-led endeavour to promote higher education through open and distance learning modes, our institution was shaped by the guiding principle that access and quality are the cornerstones of equity. We have firmly resolved to uphold the highest standards of education, setting the benchmark and charting the course.

The courses offered by the Sreenarayanaguru Open University aim to strike a quality balance, ensuring students are equipped for both personal growth and professional excellence. The University embraces the widely acclaimed "blended format," a practical framework that harmoniously integrates Self-Learning Materials, Classroom Counseling, and Virtual modes, fostering a dynamic and enriching experience for both learners and instructors.

The University aims to offer you an engaging and thought-provoking educational journey. The postgraduate programme in Economics builds on the undergraduate programme by covering more advanced theories and practical applications. The course material aims to spark learners' interest by using real-life examples and combining academic content with empirical evidence, making it relevant and unique. The Self-Learning Material has been meticulously crafted, incorporating relevant examples to facilitate better comprehension.

Rest assured, the university's student support services will be at your disposal throughout your academic journey, readily available to address any concerns or grievances you may encounter. We encourage you to reach out to us freely regarding any matter about your academic programme. It is our sincere wish that you achieve the utmost success.

Regards,
Dr. Jagathy Raj V.P.

29-04-2025

Contents

1

BLOCK

Linear Algebra and Matrix

1

UNIT

Matrix Operations

Learning Outcomes

After going through the unit, the learner will be able to:

- understand the concept of matrix inversion
- solve system of linear equations using matrix
- application of inverse of a matrix in real life

Background

Matrices form a fundamental concept in the field of linear algebra, which is a branch of mathematics dealing with linear equations, vector spaces, and transformations. One of the most important aspects is a good understanding of basic algebraic operations. For instance, being comfortable with manipulating expressions involving variables, solving equations, and comprehending how numbers interact through addition, subtraction, multiplication, and division is important.

To study matrices, familiarity with vectors and vector spaces is crucial. A vector is a mathematical entity that possesses both magnitude and direction. For instance, in a two-dimensional space, a vector can depict a displacement from point A to point B. Understanding the principles of vector addition, scaling, and analysis forms the foundation for comprehending matrix operations.

To understand the inverse of a matrix and use it to solve systems of equations, a solid foundation in basic matrix operations is essential, including addition, multiplication, and the concept of an identity matrix.

Keywords

Rank, Cramer's Rule, Inverse method

Discussion

1.1.1 Matrix Operations

Matrix operations play a crucial role in simplifying and analysing complex relationships between variables across various sectors. Key operations like finding the inverse of a matrix, determining the rank of a matrix, and using Cramer's Rule for solving simultaneous equations are powerful tools for economic modelling, forecasting, and optimizing systems.

The inverse of a 3×3 matrix allows us to solve systems of linear equations in a compact form. In economics, this can be applied to input-output models that examine how production in one sector affects other sectors. For example, in a closed economy model, we may have a matrix representing interdependencies among three sectors (e.g., agriculture, industry, and services). By calculating the inverse of this matrix, economists can determine the direct and indirect output required in each sector to meet final demand.

The rank of a matrix is a measure of its independence and can help to determine whether a system of equations has a unique solution, no solution, or infinitely many solutions. Matrix rank is crucial when dealing with data sets in regression analysis or multivariate analysis, as it indicates whether the variables are sufficiently independent to yield meaningful results.

Cramer's Rule provides a straightforward method to solve small systems of linear equations, particularly those represented by 2×2 or 3×3 matrices. Cramer's Rule can be applied in equilibrium analysis or to solve systems where demand and supply equations intersect.

Using the inverse of a matrix to solve simultaneous equations for a system $AX = B$ is widely applicable in economics, especially in linear programming and optimization. This method is essential when modeling resource allocation across industries, as it can help determine the optimal production levels required to maximize output or minimize costs.

1.1.1.1 Matrix

A system of mn-numbers (real or complex) arranged in the form of an ordered set of m horizontal lines (called rows) and n vertical lines (called columns) is an $m \times n$ matrix (to be read as $m\ by\ n$ matrix).

We write general form of $m \times n$ matrix as $\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix}$

1.1.2 Inverse of 3×3 matrix

The inverse of 3×3 matrix is a matrix that, when multiplied by the original matrix, yields the identity matrix. For a given matrix A, the inverse (if it exists) is denoted as A^{-1} and is defined only if A is invertible or non-singular, which means its determinant is non-zero.

Steps to find the inverse of a 3×3 matrix

Let the 3×3 matrix be $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$

1. Find the determinant of the matrix.

Let $A = |A| = a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$

If $|A| = 0$, the matrix is singular and the inverse does not exist.

2. Find Co factors

$C_{11} = co\ factor\ of\ a_{11} = (-1)^{1+1}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$

$C_{12} = co\ factor\ of\ a_{12} = (-1)^{1+2}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$

$C_{13} = co\ factor\ of\ a_{13} = (-1)^{1+3}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$

$C_{21} = co\ factor\ of\ a_{21} = (-1)^{2+1}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix}$

$C_{22} = co\ factor\ of\ a_{22} = (-1)^{2+2}\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$

$C_{23} = co\ factor\ of\ a_{23} = (-1)^{2+3}\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}$

$C_{31} = co\ factor\ of\ a_{31} = (-1)^{3+1}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$

$$C_{32} = co\ factor\ of\ a_{32} = (-1)^{3+2} \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}$$

$$C_{33} = co\ factor\ of\ a_{33} = (-1)^{3+3} \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

3. Form Adjoint Matrix $Adj(A)$

$$Adj(A) = \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix}$$

4. Inverse of the matrix $A = \frac{1}{|A|}\ Adj(A)$

Illustration 1.1.1

Find the inverse of the matrix $\begin{bmatrix} 2 & 3 & 4 \\ 4 & 3 & 1 \\ 1 & 2 & 4 \end{bmatrix}$

Solution

$$|A| = 2(12-2) - 3(16-1) + 4(8-3) = 2 \times 10 - 3 \times 15 + 4 \times 5$$
$$= 20 - 45 + 20 = -5$$

$$C_{11} = +\begin{vmatrix} 3 & 1 \\ 2 & 4 \end{vmatrix} = 12 - 2 = 10$$

$$C_{12} = -\begin{vmatrix} 4 & 1 \\ 1 & 4 \end{vmatrix} = -(16-1) = -15$$

$$C_{13} = +\begin{vmatrix} 4 & 3 \\ 1 & 2 \end{vmatrix} = 8 - 3 = 5$$

$$C_{21} = -\begin{vmatrix} 3 & 4 \\ 2 & 4 \end{vmatrix} = -(12-8) = -4$$

$$C_{22} = +\begin{vmatrix} 2 & 4 \\ 1 & 4 \end{vmatrix} = 8 - 4 = 4$$

$$C_{23} = -\begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix} = -(4-3) = -1$$

$$C_{31} = +\begin{vmatrix} 3 & 4 \\ 3 & 1 \end{vmatrix} = 3 - 12 = -9$$

$$C_{32} = -\begin{vmatrix} 2 & 4 \\ 4 & 1 \end{vmatrix} = -(2-16) = 14$$

$$C_{33} = +\begin{vmatrix} 2 & 3 \\ 4 & 3 \end{vmatrix} = 6 - 12 = -6$$

$$Adj(A) = \begin{bmatrix} 10 & -4 & -9 \\ -15 & 4 & 14 \\ 5 & -1 & -6 \end{bmatrix}$$

Inverse of A = $-\frac{1}{5}\begin{bmatrix} 10 & -4 & -9 \\ -15 & 4 & 14 \\ 5 & -1 & -6 \end{bmatrix}$

Illustration 1.1.2

Find the inverse of the matrix $\begin{bmatrix} 1 & 1 & 3 \\ 1 & 3 & -3 \\ -2 & -4 & -4 \end{bmatrix}$

Solution

$$|A| = 1(-12 - 12) - 1(-4 - 6) + 3(-4 + 6) = 1 \times -24 - 1 \times -10 + 3 \times 2$$

$$= -24 + 10 + 6 = -8$$

$$C_{11} = +\begin{vmatrix} 3 & -3 \\ -4 & -4 \end{vmatrix} = -12 - 12 = -24$$

$$C_{12} = -\begin{vmatrix} 1 & -3 \\ -2 & -4 \end{vmatrix} = -(-4 - 6) = 10$$

$$C_{13} = +\begin{vmatrix} 1 & 3 \\ -2 & -4 \end{vmatrix} = -4 + 6 = 2$$

$$C_{21} = -\begin{vmatrix} 1 & 3 \\ -4 & -4 \end{vmatrix} = -(-4 + 12) = -8$$

$$C_{22} = +\begin{vmatrix} 1 & 3 \\ -2 & -4 \end{vmatrix} = -4 + 6 = 2$$

$$C_{23} = -\begin{vmatrix} 1 & 1 \\ -2 & -4 \end{vmatrix} = -(-4 + 2) = 2$$

$$C_{31} = +\begin{vmatrix} 1 & 3 \\ 3 & -3 \end{vmatrix} = -3 - 9 = -12$$

$$C_{32} = -\begin{vmatrix} 1 & 3 \\ 1 & -3 \end{vmatrix} = -(-3 - 3) = 6$$

$$C_{33} = +\begin{vmatrix} 1 & 1 \\ 1 & 3 \end{vmatrix} = 3 - 1 = 2$$

$$Adj(A) = \begin{bmatrix} -24 & -8 & -12 \\ 10 & 2 & 6 \\ 2 & 2 & 2 \end{bmatrix}$$

Inverse of A = $-\frac{1}{8}\begin{bmatrix} -24 & -8 & -12 \\ 10 & 2 & 6 \\ 2 & 2 & 2 \end{bmatrix}$

Illustration 1.1.3

Find the inverse of the matrix $\begin{bmatrix} 2 & 4 & 3 \\ 0 & 1 & 1 \\ 2 & 2 & -1 \end{bmatrix}$

Solution

$$|A| = 2(-1-2) - 4(0-2) + 3(0-2) = 2 \times -3 - 4 \times -2 + 3 \times -2$$

$$= -6 + 8 - 6 = -4$$

$$C_{11} = +\begin{vmatrix} 1 & 1 \\ 2 & -1 \end{vmatrix} = -1 - 2 = -3$$

$$C_{12} = -\begin{vmatrix} 0 & 1 \\ 2 & -1 \end{vmatrix} = -(0-2) = 2$$

$$C_{13} = +\begin{vmatrix} 0 & 1 \\ 2 & 2 \end{vmatrix} = 0 - 2 = -2$$

$$C_{21} = -\begin{vmatrix} 4 & 3 \\ 2 & -1 \end{vmatrix} = -(-4-6) = 10$$

$$C_{22} = +\begin{vmatrix} 2 & 4 \\ 2 & -1 \end{vmatrix} = -2 - 6 = -8$$

$$C_{23} = -\begin{vmatrix} 2 & 4 \\ 2 & 2 \end{vmatrix} = -(4-8) = 4$$

$$C_{31} = +\begin{vmatrix} 4 & 3 \\ 1 & 1 \end{vmatrix} = 4 - 3 = 1$$

$$C_{32} = -\begin{vmatrix} 2 & 3 \\ 0 & 1 \end{vmatrix} = -(2-0) = -2$$

$$C_{33} = +\begin{vmatrix} 2 & 4 \\ 0 & 1 \end{vmatrix} = 2$$

$$Adj(A) = \begin{bmatrix} -3 & 10 & -1 \\ 2 & -8 & -2 \\ -2 & 4 & 2 \end{bmatrix}$$

Inverse of A = $-\frac{1}{4}\begin{bmatrix} -3 & 10 & -1 \\ 2 & -8 & -2 \\ -2 & 4 & 2 \end{bmatrix}$

Illustration 1.1.4

Find the inverse of the matrix $\begin{bmatrix} -1 & -2 & -2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{bmatrix}$

Solution

$$|A| = -1(1-4) + 2(2+4) - 2(-4-2) = -1 \times -3 + 2 \times 6 - 2 \times -6$$

$$= 3 + 12 + 12 = 27$$

$C_{11} = +\begin{vmatrix} 1 & -2 \\ -2 & 1 \end{vmatrix} = 1 - 4 = -3$

$C_{12} = -\begin{vmatrix} 2 & -2 \\ 2 & 1 \end{vmatrix} = -(2 + 4) = -6$

$C_{13} = +\begin{vmatrix} 2 & 1 \\ 2 & -2 \end{vmatrix} = -4 - 2 = -6$

$C_{21} = -\begin{vmatrix} -2 & -2 \\ -2 & 1 \end{vmatrix} = -(-2 - 4) = 6$

$C_{22} = +\begin{vmatrix} -1 & -2 \\ 2 & 1 \end{vmatrix} = -1 + 4 = 3$

$C_{23} = -\begin{vmatrix} -1 & -2 \\ 2 & -2 \end{vmatrix} = -(2 + 4) = -6$

$C_{31} = +\begin{vmatrix} -2 & -2 \\ 1 & -2 \end{vmatrix} = 4 + 2 = 6$

$C_{32} = -\begin{vmatrix} -1 & -2 \\ 2 & -2 \end{vmatrix} = -(2 + 4) = -6$

$C_{33} = +\begin{vmatrix} -1 & -2 \\ 2 & 1 \end{vmatrix} = -1 + 4 = 3$

$$Adj(A) = \begin{bmatrix} -3 & 6 & 6 \\ -6 & 3 & -6 \\ -6 & -6 & 3 \end{bmatrix}$$

Inverse of A = $\frac{1}{27}\begin{bmatrix} -3 & 6 & 6 \\ -6 & 3 & -6 \\ -6 & -6 & 3 \end{bmatrix}$

Illustration 1.1.5

Find the inverse of the matrix $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 5 \\ 3 & 5 & 12 \end{bmatrix}$

Solution

$$|A| = 1(36 - 25) - 2(24 - 15) + 3(10 - 9) = 1 \times 11 - 2 \times 9 + 3 \times 1$$

$$= 11 - 18 + 3 = -4$$

$C_{11} = +\begin{vmatrix} 3 & 5 \\ 5 & 12 \end{vmatrix} = 36 - 25 = 11$

$C_{12} = -\begin{vmatrix} 2 & 5 \\ 3 & 12 \end{vmatrix} = -(24 - 15) = -9$

$C_{13} = +\begin{vmatrix} 2 & 3 \\ 3 & 5 \end{vmatrix} = 10 - 9 = 1$

$$C_{21} = -\begin{vmatrix} 2 & 3 \\ 5 & 12 \end{vmatrix} = -(24-15) = -9$$

$$C_{22} = +\begin{vmatrix} 1 & 3 \\ 3 & 12 \end{vmatrix} = 12-9 = 3$$

$$C_{23} = -\begin{vmatrix} 1 & 2 \\ 3 & 5 \end{vmatrix} = -(5-6) = 1$$

$$C_{31} = +\begin{vmatrix} 2 & 3 \\ 3 & 5 \end{vmatrix} = 10-9 = 1$$

$$C_{32} = -\begin{vmatrix} 1 & 3 \\ 2 & 5 \end{vmatrix} = -(5-6) = 1$$

$$C_{33} = +\begin{vmatrix} 1 & 2 \\ 2 & 3 \end{vmatrix} = 3-4 = -1$$

$$Adj(A) = \begin{bmatrix} 11 & -9 & 1 \\ -9 & 3 & -1 \\ 1 & -1 & -1 \end{bmatrix}$$

Inverse of A $= -\frac{1}{4}\begin{bmatrix} 11 & -9 & 1 \\ -9 & 3 & -1 \\ 1 & -1 & -1 \end{bmatrix}$

1.1.3 Rank of a Matrix

Let A be an $m \times n$ matrix. Rank of a matrix is the order of the submatrix of A whose determinant $\neq 0$. Submatrix is a matrix obtained by eliminating rows or column of a matrix.

i.e. It is the maximum number of independent columns of a matrix.

In order to find the rank of a matrix we have to apply elementary transformations and reduce the matrix into echelon form. Rank of the matrix in the echelon form is the number of non zero rows.

Elementary transformation

1. Interchange of two rows, $R_i \leftrightarrow R_j$.

2. Multiplication of any row by a non zero number, $R_i \to k\ R_i$.

3. Addition of each row by a constant multiple of other row, $R_i \to k\ R_i + R_j$.

Echelon Form

1. Any row that consists entirely of zero must be at the bottom of the matrix ..

2. The number of zero rows before the first non-zero element is in increasing order.

3. The rank of the matrix in the echelon form is the number of non-zero rows.

Example, $\begin{bmatrix} 3 & -2 & 1 \\ 0 & -7 & 2 \\ 0 & 0 & -4 \end{bmatrix}$ Rank = 3, $\begin{bmatrix} 3 & -2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ Rank = 1.

Illustration 1.1.6

Find the rank of $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

Solution

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & -1 \\ 0 & 1 & 1 \end{bmatrix} \qquad R_2 \to R_2 + (-2)R_1, R_3 \to R_3 + (-1)R_1$$

$$\sim \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \qquad R_3 \to R_3 + R_2$$

$Rank = 2$

Illustration 1.1.7

Find the rank of $\begin{bmatrix} 1 & 2 & -2 \\ -1 & 3 & 0 \\ 0 & -2 & 1 \end{bmatrix}$

Solution

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 5 & -2 \\ 0 & -2 & 1 \end{bmatrix} \qquad R_2 \to R_2 + R_1,$$

$$\sim \begin{bmatrix} 1 & 2 & 1 \\ 0 & 5 & -2 \\ 0 & 0 & 1 \end{bmatrix} \qquad R_3 \to 5R_3 + 2R_2$$

$Rank = 3$

Illustration 1.1.8

Find the rank of $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ -3 & -6 & -9 \end{bmatrix}$

Solution

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad R_2 \to R_2 + (-2)R_1, \qquad R_3 \to R_3 + (3)R_1$$

$Rank = 1$

Illustration 1.1.9

Find the rank of $\begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & -1 & 2 & -1 \\ 3 & 1 & 0 & 1 \end{bmatrix}$

Solution

$$\sim \begin{bmatrix} 1 & 1 & -1 & 1 \\ 0 & -2 & 3 & -2 \\ 0 & -2 & 3 & -2 \end{bmatrix} \qquad R_2 \to R_2 + (-1)R_1, R_3 \to R_3 + (-3)R_1$$

$$\sim \begin{bmatrix} 1 & 1 & -1 & 1 \\ 0 & -2 & 3 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad R_3 \to R_3 + (-1)R_2$$

$Rank = 2$

Illustration 1.1.10

Find the rank of $\begin{bmatrix} 3 & 1 & 2 & 0 \\ 1 & 0 & -1 & 0 \\ 2 & 1 & 3 & 0 \end{bmatrix}$

Solution

$$\sim \begin{bmatrix} 3 & 1 & 2 & 0 \\ 0 & 1 & 5 & 0 \\ 0 & 1 & 5 & 0 \end{bmatrix} \qquad R_2 \to (-3)R_2 + R_1, \;\; R_3 \to 3R_3 + (-2)R_1$$

$$\sim \begin{bmatrix} 3 & 1 & 2 & 0 \\ 0 & 1 & 5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad R_3 \to R_3 + (-1)R_2$$

$Rank = 2$

1.1.4 Solution of System of equations

Consider the system of equations

$$a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n = b_2$$

$$a_{31}x_1 + a_{32}x_2 + \cdots a_{3n}x_n = b_3$$

…………………………………...

$$a_{m1}x_1 + a_{m2}x_2 + \cdots a_{mn}x_n = b_m$$

$$x_1, x_2, \ldots x_n \geq 0$$

The numbers a_{ij} are called coefficients and b_i is constants, $i = 1,2 \ldots m, j = 1,2 \ldots n.$

When all b_i are not zero, i.e. at least one $b_i \neq 0$, then the system is non homogeneous.

If all $b_i = 0$, then the system is homogeneous.

The system of equations can be represented in a matrix form $AX = B$

where $X = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{bmatrix}$ $B = \begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ b_m \end{bmatrix}$ and $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$

1.1.5 Solution of Linear System of Equations by Cramer's Rule

Let us discuss a determinant-based method for solving systems of equations. This technique, known as Cramer's Rule, dates back to the middle of the 18th century and is named after its inventor, Swiss mathematician Gabriel Cramer (1704-1752), who introduced it in 1750 in "Introduction to the analysis of algebraic curved lines", a geometry book. Cramer's Rule is a viable and efficient method for solving systems with an arbitrary number of unknowns, provided that the number of equations equals the number of unknowns.

Let us see how the Cramer's Rule works using the following example.

Consider a set of simultaneous equations say,

$2x + 3y = 1, \; 3x + y = 5$

Step 1

Write the equations in the matrix form as follows,

$\begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$ that is AX = B.

where, $A = \begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix}$, $X = \begin{bmatrix} x \\ y \end{bmatrix}$, $B = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$

Step 2

Obtain A_1 and A_2.

A_1 is obtained by replacing first column of A by B. That is,

$A_1 = \begin{bmatrix} 1 & 3 \\ 5 & 1 \end{bmatrix}$

A_2 is obtained by replacing second column of A by B. That is,

$A_2 = \begin{bmatrix} 2 & 1 \\ 3 & 5 \end{bmatrix}$

Step 3

Find the values of the determinants $A, A_1 \; and \; A_2$.

That is,

$|A| = \begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix}$ = 2 – 9 = -7.

$|A_1| = \begin{vmatrix} 1 & 3 \\ 5 & 1 \end{vmatrix}$ = 1 – 15 = -14.

$|A_2| = \begin{vmatrix} 2 & 1 \\ 3 & 5 \end{vmatrix}$ = 10 – 3 = 7.

Step 4

Find the values of the unknowns using the formula,

$$x = \frac{|A_1|}{|A|}, \qquad y = \frac{|A_2|}{|A|}$$

In our example,

$$x = \frac{|A_1|}{|A|} = \frac{-14}{-7} = 2$$

$$y = \frac{|A_2|}{|A|} \frac{7}{-7} = 1$$

Therefore, the value of $x = 2 \; and \; y = 1$

Illustration 1.1.11

Solve the linear equations using Cramer's Rule.

$2x + 3y = 7$

$4x - 3y = 5$

Solution

First of all, let us write the equations in the matrix form as follows,

$\begin{bmatrix} 2 & 3 \\ 4 & -3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ 5 \end{bmatrix}$ that is AX = B.

A = $\begin{bmatrix} 2 & 3 \\ 4 & -3 \end{bmatrix}$, X = $\begin{bmatrix} x \\ y \end{bmatrix}$, B = $\begin{bmatrix} 7 \\ 5 \end{bmatrix}$

$|A| = \begin{vmatrix} 2 & 3 \\ 4 & -3 \end{vmatrix} = (2 \times -3) - (4 \times 3) = -6 - 12 = -18$

$A_1 = \begin{bmatrix} 7 & 3 \\ 5 & -3 \end{bmatrix}$

$$|A_1| = \begin{vmatrix} 7 & 3 \\ 5 & -3 \end{vmatrix} = (7 \times -3) - (3 \times 5) = -21 - 15 = -36$$

$$A_2 = \begin{bmatrix} 2 & 7 \\ 4 & 5 \end{bmatrix}$$

$$|A_2| = \begin{vmatrix} 2 & 7 \\ 4 & 5 \end{vmatrix} = (2 \times 5) - (7 \times 4) = 10 - 28 = -18$$

Therefore,

$$x = \frac{|A_1|}{|A|} = \frac{-36}{-18} = 2,$$

$$y = \frac{|A_2|}{|A|} = \frac{-18}{-18} = 1$$

Thus, $x = 2 \; and \; y = 1$

Illustration 1.1.12

Solve the linear equations using Cramer's Rule.

$x + y + z = 11, \; 2x - 6y - z = 0 \; and \; 3x + 4y + 2z = 0.$

Solution

The given liner equations are,

$x + y + z = 11, \; 2x - 6y - z = 0 \; and \; 3x + 4y + 2z = 0.$

Let us write the given equations in the matrix form as follows,

$\begin{bmatrix} 1 & 1 & 1 \\ 2 & -6 & -1 \\ 3 & 4 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 11 \\ 0 \\ 0 \end{bmatrix}$ that is $AX = B$.

Where,

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & -6 & -1 \\ 3 & 4 & 2 \end{bmatrix}, \quad X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad B = \begin{bmatrix} 11 \\ 0 \\ 0 \end{bmatrix}$$

Since $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & -6 & -1 \\ 3 & 4 & 2 \end{bmatrix}$

$$|A| = \begin{bmatrix} 1 & 1 & 1 \\ 2 & -6 & -1 \\ 3 & 4 & 2 \end{bmatrix} = 1 \begin{vmatrix} -6 & -1 \\ 4 & 2 \end{vmatrix} - 1 \begin{vmatrix} 2 & -1 \\ 3 & 2 \end{vmatrix} + 1 \begin{vmatrix} 2 & -6 \\ 3 & 4 \end{vmatrix}$$

$$= 1(-12 + 4) - 1(4 + 3) + 1(8 - (-18))$$

$$= 1(-8) - 1(7) + 1(26)$$

$= -8 - 7 + 26 = 11.$

$$A_1 = \begin{bmatrix} 11 & 1 & 1 \\ 0 & -6 & -1 \\ 0 & 4 & 2 \end{bmatrix}$$

$$|A_1| = \begin{bmatrix} 11 & 1 & 1 \\ 0 & -6 & -1 \\ 0 & 4 & 2 \end{bmatrix} = 11 \begin{vmatrix} -6 & -1 \\ 4 & 2 \end{vmatrix} - 1 \begin{vmatrix} 0 & -1 \\ 0 & 2 \end{vmatrix} + \begin{vmatrix} 0 & -6 \\ 0 & 4 \end{vmatrix}$$

$= 11\,(-12 + 4) - 1\,(0 - 0) + (0 - 0)$

$= 11\,(-8) - 1\,(0) + (0)$

$= -88.$

$$A_2 = \begin{bmatrix} 1 & 11 & 1 \\ 2 & 0 & -1 \\ 3 & 0 & 2 \end{bmatrix}$$

$$|A_2| = \begin{bmatrix} 1 & 11 & 1 \\ 2 & 0 & -1 \\ 3 & 0 & 2 \end{bmatrix} = 1 \begin{vmatrix} 0 & -1 \\ 0 & 2 \end{vmatrix} - 11 \begin{vmatrix} 2 & -1 \\ 3 & 2 \end{vmatrix} + 1 \begin{vmatrix} 2 & 0 \\ 3 & 0 \end{vmatrix}$$

$= 1\,(0 - 0) - 11\,(4 + 3) + 1\,(0 - 0)$

$= -11\,(7)$

$= -77.$

$$A_3 = \begin{bmatrix} 1 & 1 & 11 \\ 2 & -6 & 0 \\ 3 & 4 & 0 \end{bmatrix}$$

$$|A_3| = \begin{bmatrix} 1 & 1 & 11 \\ 2 & -6 & 0 \\ 3 & 4 & 0 \end{bmatrix} = 1 \begin{vmatrix} -6 & 0 \\ 4 & 0 \end{vmatrix} - 1 \begin{vmatrix} 2 & 0 \\ 3 & 0 \end{vmatrix} + 11 \begin{vmatrix} 2 & -6 \\ 3 & 4 \end{vmatrix}$$

$= (0 - 0) - 1\,(0 - 0) + 11\,(8 - (-18))$

$= 11 \times 26$

$= 286$

$$x = \frac{|A_1|}{|A|} = \frac{-88}{11} = -8$$

$$y = \frac{|A_2|}{|A|} = \frac{-77}{11} = -7$$

$$z = \frac{|A_3|}{|A|} = \frac{286}{11} = 26$$

Therefore, $x = -8, y = -7, z = 26.$

Illustration 1.1.13

Solve the linear equations using Cramer's Rule.

$3x + y + z = 8, \quad x + y + z = 6 \ and \ 2x + y - z = 1$

Let us write the given equations in the matrix form as follows,

$$\begin{bmatrix} 3 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \\ 1 \end{bmatrix} \text{ that is } AX = B.$$

Where,

$$A = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & -1 \end{bmatrix}, \quad X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad B = \begin{bmatrix} 8 \\ 6 \\ 1 \end{bmatrix}$$

Since $A = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & -1 \end{bmatrix}$

$$|A| = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & -1 \end{bmatrix} = 3\begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} - 1\begin{vmatrix} 1 & 1 \\ 2 & -1 \end{vmatrix} + 1\begin{vmatrix} 1 & 1 \\ 2 & 1 \end{vmatrix}$$

$$= 3(-1-1) - 1(-1-2) + 1(1-2)$$

$$= 3(-2) - 1(-3) + 1(-1)$$

$$= -6 + 3 - 1 = -4$$

$$A_1 = \begin{bmatrix} 8 & 1 & 1 \\ 6 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}$$

$$|A_1| = \begin{bmatrix} 8 & 1 & 1 \\ 6 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix} = 8\begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} - 1\begin{vmatrix} 6 & 1 \\ 1 & -1 \end{vmatrix} + 1\begin{vmatrix} 6 & 1 \\ 1 & 1 \end{vmatrix}$$

$$= 8(-1-1) - 1(-6-1) + (6-1)$$

$$= 8(-2) - 1(-7) + (5)$$

$$= -16 + 7 + 5$$

$$= -4$$

$$A_2 = \begin{bmatrix} 3 & 8 & 1 \\ 1 & 6 & 1 \\ 2 & 1 & -1 \end{bmatrix}$$

$$|A_2| = \begin{bmatrix}3 & 8 & 1\\ 1 & 6 & 1\\ 2 & 1 & -1\end{bmatrix} = 3\begin{vmatrix}6 & 1\\ 1 & -1\end{vmatrix} - 8\begin{vmatrix}1 & 1\\ 2 & -1\end{vmatrix} + 1\begin{vmatrix}1 & 6\\ 2 & 1\end{vmatrix}$$

$$= 3\,(-6-1) - 8\,(-1-2) + 1\,(1-12)$$

$$= 3(-7) - 8\,(-3) + (-11)$$

$$= -21 + 24 - 11$$

$$= -8.$$

$$A_3 = \begin{bmatrix}3 & 1 & 8\\ 1 & 1 & 6\\ 2 & 1 & 1\end{bmatrix}$$

$$|A_3| = \begin{bmatrix}3 & 1 & 8\\ 1 & 1 & 6\\ 2 & 1 & 1\end{bmatrix} = 3\begin{vmatrix}1 & 6\\ 1 & 1\end{vmatrix} - \begin{vmatrix}1 & 6\\ 2 & 1\end{vmatrix} + 8\begin{vmatrix}1 & 1\\ 2 & 1\end{vmatrix}$$

$$= 3\,(1-6) - (1-12) + 8\,(1-2)$$

$$= 3(-5) - (-11) + 8\,(-1)$$

$$= -15 + 11 - 8$$

$$= -12$$

$$x = \frac{|A_1|}{|A|} = \frac{-4}{-4} = 1$$

$$y = \frac{|A_2|}{|A|} = \frac{-8}{-4} = 2$$

$$z = \frac{|A_3|}{|A|} = \frac{-12}{-4} = 3$$

Therefore, $x = 1,\ y = 2,\ z = 3$

1.1.6 Matrix Inverse Method

Let the given system of equation is $AX = B$ where A is the n - square non-singular matrix. X is the variable matrix and B is the constant matrix.

i.e. $A^{-1}(AX) = A^{-1}B$

$$(A^{-1}A)X = A^{-1}B$$

$$X = A^{-1}B$$

which is the required solution where A^{-1} is the inverse of A and $A^{-1} = \frac{adjA}{|A|}$.

Illustration 1.1.14

Solve the linear equations using Matrix Inverse Method

$x + y + 3z = 1, \quad x + 3y - 3z = 2 \text{ and } -2x - 4y - 4z = 2$

Solution

$$A = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 3 & -3 \\ -2 & -4 & -4 \end{bmatrix}, \quad X = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad B = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

$$|A| = 1(-12-12) - 1(-4-6) + 3(-4+6) = 1 \times -24 - 1 \times -10 + 3 \times 2$$

$$= -24 + 10 + 6 = -8$$

$$C_{11} = +\begin{vmatrix} 3 & -3 \\ -4 & -4 \end{vmatrix} = -12 - 12 = -24$$

$$C_{12} = -\begin{vmatrix} 1 & -3 \\ -2 & -4 \end{vmatrix} = -(-4-6) = 10$$

$$C_{13} = +\begin{vmatrix} 1 & 3 \\ -2 & -4 \end{vmatrix} = -4 + 6 = 2$$

$$C_{21} = -\begin{vmatrix} 1 & 3 \\ -4 & -4 \end{vmatrix} = -(-4+12) = -8$$

$$C_{22} = +\begin{vmatrix} 1 & 3 \\ -2 & -4 \end{vmatrix} = -4 + 6 = 2$$

$$C_{23} = -\begin{vmatrix} 1 & 1 \\ -2 & -4 \end{vmatrix} = -(-4+2) = 2$$

$$C_{31} = +\begin{vmatrix} 1 & 3 \\ 3 & -3 \end{vmatrix} = -3 - 9 = -12$$

$$C_{32} = -\begin{vmatrix} 1 & 3 \\ 1 & -3 \end{vmatrix} = -(-3-3) = 6$$

$$C_{33} = +\begin{vmatrix} 1 & 1 \\ 1 & 3 \end{vmatrix} = 3 - 1 = 2$$

$$Adj(A) = \begin{bmatrix} -24 & -8 & -12 \\ 10 & 2 & 6 \\ 2 & 2 & 2 \end{bmatrix}$$

Inverse of A $= -\frac{1}{8}\begin{bmatrix} -24 & -8 & -12 \\ 10 & 2 & 6 \\ 2 & 2 & 2 \end{bmatrix}$

$$X = A^{-1}B = -\frac{1}{8}\begin{bmatrix} -24 & -8 & -12 \\ 10 & 2 & 6 \\ 2 & 2 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

$$= -\frac{1}{8}\begin{bmatrix} -24-16-24 \\ 10+\ 4\ +12 \\ 2+4\ +\ \ 4 \end{bmatrix}$$

$$= -\frac{1}{8}\begin{bmatrix} -64 \\ 26 \\ 10 \end{bmatrix}$$

Therefore, $x\ = \frac{64}{8} = 8,\qquad y = -\frac{26}{8} = \frac{13}{4},\qquad z = -\frac{10}{8} = -\frac{5}{4}$

Illustration 1.1.15

Solve the linear equations using Matrix Inverse Method

$-x - 2y - 2\,z\ = 1,\ \ 2x\ +\ y - 2\,z\ \ =\ 1\ \ and\ \ 2x - 2\,y + \ z\ =\ 1$

Solution

$A = \begin{bmatrix} -1 & -2 & -2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{bmatrix},\ X = \begin{bmatrix} x \\ y \\ z \end{bmatrix},\ B = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

$$|A| = -1(1-4) + 2(2+4) - 2(-4-2) = -1 \times -3 + 2 \times 6 - 2 \times -6$$

$$= 3 + 12 + 12 = 27$$

$C_{11} = +\begin{vmatrix} 1 & -2 \\ -2 & 1 \end{vmatrix} = 1 - 4 = -3$

$C_{12} = -\begin{vmatrix} 2 & -2 \\ 2 & 1 \end{vmatrix} = -(2+4) = -6$

$C_{13} = +\begin{vmatrix} 2 & 1 \\ 2 & -2 \end{vmatrix} = -4 - 2 = -6$

$C_{21} = -\begin{vmatrix} -2 & -2 \\ -2 & 1 \end{vmatrix} = -(-2-4) = 6$

$C_{22} = +\begin{vmatrix} -1 & -2 \\ 2 & 1 \end{vmatrix} = -1 + 4 = 3$

$C_{23} = -\begin{vmatrix} -1 & -2 \\ 2 & -2 \end{vmatrix} = -(2+4) = -6$

$C_{31} = +\begin{vmatrix} -2 & -2 \\ 1 & -2 \end{vmatrix} = 4 + 2 = 6$

$C_{32} = -\begin{vmatrix} -1 & -2 \\ 2 & -2 \end{vmatrix} = -(2+4) = -6$

$C_{33} = +\begin{vmatrix} -1 & -2 \\ 2 & 1 \end{vmatrix} = -1 + 4 = 3$

$$Adj(A) = \begin{bmatrix} -3 & 6 & 6 \\ -6 & 3 & -6 \\ -6 & -6 & 3 \end{bmatrix}$$

Inverse of A = $\frac{1}{27}\begin{bmatrix} -3 & 6 & 6 \\ -6 & 3 & -6 \\ -6 & -6 & 3 \end{bmatrix}$

$$X = A^{-1}B = \frac{1}{27}\begin{bmatrix} -3 & 6 & 6 \\ -6 & 3 & -6 \\ -6 & -6 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \frac{1}{27}\begin{bmatrix} -3+6+6 \\ -6+3-6 \\ -6-6+3 \end{bmatrix}$$

$$= \frac{1}{27}\begin{bmatrix} 9 \\ -9 \\ -9 \end{bmatrix}$$

Therefore, $x = \frac{1}{3}$, $y = -\frac{1}{3}$, $z = -\frac{1}{3}$.

Summarised Overview

Matrix operations are essential tools in economics for modeling, analysing, and optimising complex systems. Matrix Inverse is used to solve systems of linear equations, particularly in input-output models to determine sector interdependencies (e.g., agriculture, industry, services). The inverse exists only for non-singular matrices (determinant ≠ 0). Matrix Rank indicates the number of linearly independent rows or columns, helping assess whether a system has a unique, infinite, or no solution. Crucial for regression and multivariate analysis. Cramer's Rule is a determinant-based method to solve small linear systems (e.g., 2×2 or 3×3), useful in equilibrium analysis and demand-supply intersections. Matrix Inverse Method solves systems AX=B via X=A−1B, applied in resource allocation and optimization problems.

Inverse of 3×3 Matrices is calculated using determinants and adjugates (steps: compute determinant, cofactors, adjoint, then $A^{-1} = \frac{1}{|A|}$ Adj (A). Examples provided for singular or non-singular cases. Rank Determination is achieved via row reduction to echelon form (rank = non-zero rows). Examples illustrate dependency checks. Cramer's Rule solves systems by replacing matrix columns with constants and computing determinants (e.g., $x = |A_1|/|A|$) Matrix Inverse Method is efficient for larger systems, requiring A^{-1} to compute X directly. Applied in economic models like sectoral output optimization. These methods underpin economic modeling, forecasting, and system optimization by simplifying complex relationships into solvable algebraic frameworks.

Assignments

1. Solve the system of equations by Cramer's rule

   $5x - 2y + 3z = 16, \ 2x + 3y - 5z = 2 \ and \ 4x - 5y + 6z = 7.$

2. Solve the system of equations by Cramer's rule

   $11x - y - z = 31, \ -x + 6y - 2z = 26 \ and - x - 2y + 7z = 24.$

3. Find the rank of the matrix $\begin{bmatrix} 8 & 3 & 2 \\ 6 & 4 & 7 \\ 5 & 1 & 3 \end{bmatrix}$

4. Find the rank of the matrix $\begin{bmatrix} -3 & 6 & 2 \\ 1 & 5 & 4 \\ 4 & -8 & 2 \end{bmatrix}$

5. Solve the system of equations by Inverse Method

   $4x + 3y = 28, \ 2x + 5y = 42$

6. Solve the system of equations by Inverse Method

   $2x + 4y - 3z = 12, \ 3x - 5y + 2z = 13 \ and - x + 3y + 2z = 17.$

References

1. Yamane, Taro. (2012). *Mathematics for Economists: An Elementary Survey.* New Delhi: Prentice Hall of India.

2. Chiang, A.C. (2008), *Fundamental Methods of Mathematical Economics*, McGraw Hill, New York.

3. Gujarathi , D.&Sangeetha, N. (2007). *Basic Econometrics* (4thed) New Delhi: McGraw Hill

Suggested Readings

1. Sydsaeter, K., Hammond, P., Seierstad, A., & Strom, A. (2008*). Further Mathematics for Economic Analysis* (2nd ed.). Pearson Education.

2. Simon, C. P., & Blume, L. (1994). *Mathematics for Economists*. W. W. Norton & Company.

3. Hoffman, K., & Kunze, R. (1971). *Linear Algebra* (2nd ed.). Prentice-Hall.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

2 UNIT Matrix Applications in Economic Modeling

Learning Outcomes

After completing this unit, the learner will be able to:

- examine the practical applications of matrices in economic analysis
- differentiate between the static and dynamic input-output models
- apply the Hawkin-Simon condition to determine the stability of economic systems

Background

In economics, mathematical tools and techniques play a crucial role in simplifying complex relationships and making economic analysis more precise and practical. One such powerful mathematical tool is the matrix, which is widely used in various economic models to represent and analyse interdependencies among different sectors of an economy. Among the most notable applications of matrices in economics is the Input-Output analysis, originally developed by the Nobel Laureate Wassily W Leontief. Input-output analysis is a quantitative economic technique that models how different industries interact with each other within an economy. It captures the relationship between inputs (resources or goods used for production) and outputs (final goods and services) across different sectors. This model allows economists and policymakers to examine how changes in one sector impact other sectors, helping with better planning and forecasting.

In this unit, we will discuss on how matrices are used specifically in input-output models, particularly focusing on different forms of the model - static vs dynamic and open vs closed models. A fundamental condition that ensures the feasibility and stability of input-output models is Hawkins-Simon condition, which provides criteria to determine whether the system of equations representing the input-output structure has a viable solution. It is essential to ensure that all sectors in the model can produce a non-negative level of output given the inter-sectoral dependencies.

Let us now examine in detail how matrices are used in input-output analysis, the differences between the various forms of the model, and the role of the Hawkins-Simon condition in determining the viability of such models.

Keywords

Input-output model, Technical Coefficient Matrix, Static, Dynamic, Open and Closed Input-Output Model, Hawkins - Simon Viability Condition

Discussion

1.2.1 Range

Matrices have proved their usefulness in quantitative analysis of managerial decisions in several disciplines like marketing, finance, production, economics etc. Many quantitative methods such as linear programming, game theory, Markov models, input-output models and some statistical models have matrix algebra as their underlying theoretical base. All these models are built by establishing a system of linear equations, which represent the problem to be solved. Once the system of equations is represented in matrix form, they can be solved easily and quickly.

Matrix algebra is a very powerful mathematical tool to deal with the problem of solving a system of simultaneous linear equations. In most economic problems, the behaviour of economic variables are normally assumed to be linear although non-linear relations exist in practice. Even within the restrictive assumptions of linearity, we are to find out the equilibrium values of the endogenous variables in a system of linear simultaneous equations. In real life, we do not have to deal with a single magnitude but with a set of magnitudes. In analysis, we treat such a set of magnitudes as a single entity or a single object of thought, abstracted by numbers. Let us consider a problem where we treat a set as single entity and a number of such sets as a single system. We will then examine how the four fundamental operations of arithmetic can be suitably defined to deal with such a system.

Suppose there are three friends : Sanu (S), Deepak (D) and Ajay (A) in a hostel and that

S has a set of 4 pants, 4 shirts, 3 Bush-shirts and 2 Ties

D has a set of 6 pants, 6 shirts, No Bush shirts, and 3 Ties

A has a set of 7 pants, 9 shirts, 6 Bush-shirts and No Tie

We can arrange this data in the following convenient system:

$$\begin{array}{cccccc}
\text{Friends} & \text{Pants} & \text{Shirts} & \text{Bush-shirts} & \text{Ties} & \\
\downarrow & \text{(P)} & \text{(S)} & \text{(B)} & \text{(T)} & \\
S \rightarrow & 4 & 4 & 3 & 2 & \rightarrow 1^{st}\ \text{Row} \\
D \rightarrow & 6 & 6 & 0 & 3 & \rightarrow 2^{nd}\ \text{Row} \\
A \rightarrow & 7 & 9 & 6 & 0 & \rightarrow 3^{rd}\ \text{Row} \\
 & \downarrow & \downarrow & \downarrow & \downarrow & \\
 & 1^{st} & 2^{nd} & 3^{rd} & 4^{th} & \\
 & \text{Col.} & \text{Col.} & \text{Col.} & \text{Col.} &
\end{array}$$

This system comprises 3 rows and 4 columns. We have written belongings of Sanu in the first row, belongings of Deepak in the second row and the belongings of Ajay in the third row. Clearly, therefore, 1st column gives us the total number of pants that S, D and A together have; while second and third and fourth columns enumerate respectively the number of shirts, Bush Shirts and Ties that three friends together have in the hostel.

Numbers written in such a particular form of rows and columns and enclosed by square brackets - [ ] or large parentheses - ( ).

A matrix is defined as a rectangular array of elements arranged in rows and columns.

Consider the matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots\cdots & a_{1n} \\ a_{21} & a_{22} & \cdots\cdots & a_{2n} \\ \vdots & \vdots & \cdots\cdots & \vdots \\ a_{m1} & a_{m2} & \cdots\cdots & a_{mn} \end{bmatrix}$$

Where $a_{11,}\, a_{12}$ .... a_{mn} denote the numbers (elements) of the matrix. The dimension (order) of the matrix is determined by the number of rows and columns. Here, in the given matrix, there are m rows and n columns. Therefore, it is of the dimension $m\ x\ n$ (m by n). In the dimension of the given matrix the number of rows is always specified first and then the number of columns. Thus, a_{13} belongs to the 1st row and 3rd column and a_{34} belongs to the 3rd row and 4th column. In general, a_{ij} refers to the i^{th} row and j^{th} column. Boldface capital letters such as A, B, C ……….are used to denote the entire matrix. The matrix is also sometimes represented as $A = [a_{ij}]_{mxn}$. Some examples of the matrices are

$$A = \begin{bmatrix} -2 & 2 \\ 4 & 6 \end{bmatrix}_{2x2} \qquad B = \begin{bmatrix} 1 & 1 & 3 \\ 4 & 8 & -6 \end{bmatrix}_{2x3} \qquad C = \begin{bmatrix} 10 & 10 & 20 \\ 12 & 4 & 20 \\ 4 & 2 & 4 \end{bmatrix}_{3x3}$$

$$A = \begin{bmatrix} -2 & 2 \\ 4 & 6 \end{bmatrix}_{2x2} \quad B = \begin{bmatrix} 1 & 1 & 3 \\ 4 & 8 & -6 \end{bmatrix}_{2x3} \qquad C = \begin{bmatrix} 10 & 10 & 20 \\ 12 & 4 & 20 \\ 4 & 2 & 4 \end{bmatrix}_{3x3}$$

The matrix A is a 2×2 matrix because it has 2 rows and 2 columns. Similarly, the matrix B is a 2×3 matrix while matrix C is a 3×3 matrix.

In general, governments and economists use matrices to understand how different sectors are interlinked. There is a sectoral interdependence analysis. Matrices help to assess the ripple effects of policy changes or external shocks on all sectors and also the regional planners use matrix-based models to allocate resources optimally across districts or states. It helps to estimate how changes in demand for a product or service affect employment and income across the economy.

1.2.2 Input - Output Models

One of the very important areas of economic study for matrix algebra is most frequently and rigorously used in the input output analysis first advocated by professor Wassily W Leontief in the year 1951. The input - output model is based on the preposition of Walrasian general equilibrium where all the producing sectors are assumed to be inter-related. This method is based on the concept of 'economic interdependence', which means that every sector or industry of the economy is related to every other sector. That is, they are all in the dependent and inter-related. This means, any change in one sector will affect all other industries to a varying degree. This technique deals with the type of problems, one of which may be described in the following words:

"What should be the level of output of each industry with the existing technology so that the total output goal for consumer and industrial use of its product gets fully satisfied; or alternatively, what level of output of each producing sector in an economy can bring about equilibrium for its product in the economy as a whole."

The basic idea behind this problem is quite simple to understand. Since inputs of one industry are outputs of another industry and *vice versa,* ultimately their mutual relationship must lead to equilibrium between supply and demand in the economy consisting of *n* industries. For example, the output of industry one (1) is needed as an input in many other industries and perhaps for that industry itself; naturally, therefore, the total output level of industry one (1) must take account of the input requirements of all the industries in the economy. Exactly in the same way since the output of *n* enters into other industries as their 'input requirements' the total output of n^{th} industry must be one that is consistent with all input requirements so as to avoid any bottlenecks anywhere in the economy. Thus, the essence of input output analysis is that, given certain technological coefficients and final demand, each endogenous sector would find its output uniquely determined as a linear combination of multi-sector demand.

Let us suppose that an economic system consist of *n* producing sectors. In order to avoid bottlenecks in the economy, the total output of each producing sector must satisfy the total demand for its product which, in fact, would arise because:

1. Its product is being used as an intermediate product, that is, it input elsewhere in the industrial structure of production.

2. Its product is used for household consumption, capital formation, Government consumption or for export.

For example, total output of agriculture sector may be used (or demanded):

1. As an input in food or other manufacturing sector (grain for producing bread or cotton for producing cloth)

2. As a final consumption by Government or households (vegetables or grain) and/or as an export demand.

Let us assume that an economy consists of four (4) producing sectors only, and that the production of each sector is being used as an input in all the sectors and is used for final consumption.

Suppose:(i) X_1, X_2, X_3 and X_4 are total outputs of the 4 sectors.

(ii) F_1, F_2, F_3 and F_4 are the amounts of final demand, consumption, capital formation and exports for output of these sectors.

Input -Output Transaction Table

Producing sector No.	**Total output of the sector**	**Input requirements of producing sectors**				**Requirements for final uses**
		X_1	X_2	X_3	X_4	
1	**2**	**3**	**4**	**5**	**6**	**7**
1	X_1	X_{11}	X_{12}	X_{13}	X_{14}	F_1
2	X_2	X_{21}	X_{22}	X_{23}	X_{24}	F_2
3	X_3	X_{31}	X_{32}	X_{33}	X_{34}	F_3
4	X_4	X_{41}	X_{42}	X_{43}	X_{44}	F_4
Primary input (Labour)	Total primary input = L →	L_1	L_2	L_3	L_4	-

We derive two important equations from the above table:

Columns 3, 4, 5 and 6 of the above table gives us total inputs from all sectors utilized by each sector for its production. In other words, column. 3 gives the production function of sector 1 and column. 6 represents the production function of sector 4.

$$X_1 = f_1 (X_{11,} X_{21,} X_{31,} X_{41,} L_1)$$

$$X_2 = f_1 (X_{12,} X_{22,} X_{32,} X_{42,} L_2)$$

$$X_3 = f_1 (X_{13,} X_{23,} X_{33,} X_{43,} L_3)$$

$$X_4 = f_1 (X_{14,} X_{24,} X_{34,} X_{44,} L_4)$$

In general terms, if there are *'n'* number of producing sectors then the production function of sector *n* will be represented by

$$X_n = f_n (X_{1n,} X_{2n,} X_{3n} \ldots\ldots\ldots.., X_{4n,} L_n)$$

Rows of the table give us the equality between the demand and supply of each product:

$$X_1 = X_{11} + X_{12} + X_{13} + X_{14} + F_1$$

$$X_2 = X_{21} + X_{22} + X_{23} + X_{24} + F_2$$

$$X_3 = X_{31} + X_{32} + X_{33} + X_{34} + F_3$$

$$X_4 = X_{41} + X_{42} + X_{43} + X_{44} + F_4$$

$$L = L_1 + L_2 + L_3 + L_4$$

In general terms, if there are *n* producing sectors:

$$X_1 = X_{11} + X_{12} + X_{13} + \ldots\ldots\ldots\ldots X_{1n} + F_1$$

$$X_2 = X_{21} + X_{22} + X_{23} + \ldots\ldots\ldots\ldots X_{2n} + F_2$$

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

$$X_n = X_{n1} + X_{n2} + X_{n3} + \ldots\ldots\ldots\ldots X_{nn} + F_2$$

$$L = L_1 + L_2 + L_3 + L_4 + \ldots\ldots\ldots\ldots + L_n$$

That is, $X_i = \sum_{j=1}^{n} X_{ij} + F_i$ and $L = \sum_{i=1}^{n} L_i$

Here, X_i = Total output of *i*th sector

X_{ij} = Output of *i*th sector used as input in *j*th sector

F_i = Final demand for the *i*th sector

The above identity states that all the output of a particular structure could be utilised either as an input in one of the producing sectors of the economy and/or as a final demand.

1.2.3 Static Open Input-Output Analysis

The input-output model is based on the preposition of Walras general equilibrium where all the producing sectors are assumed to be inter-related.

Assumptions

The economy can be meaningfully divided into a finite number of sectors/industries:

1. Each sector/industry produces only one homogeneous commodity and so it rules out joint production.
2. Each sector/industry uses fixed input ratio or factor combination.
3. The production of each sector/industries is subject to constant returns to scale.

Based on the above assumptions, the input-output model deals with the problem of internal consistency in terms of equality of demand and supply of sectoral outputs.

The Technological Coefficient Matrix

From the assumption of fixed input requirements, we see that in order to produce one unit of j^{th} commodity, the input used of i^{th} commodity must be a fixed amount, which we denote by a_{ij}; thus $a_{ij} = (X_{ij})/X_j$. if X_j represents the total output of the j^{th} commodity/j^{th} producing sector the input requirements of i^{th} commodity will be equal to $a_{ij}X_j$ or $X_{ij} = a_{ij}X_j$.

As such we can now put the input-output transaction table in terms of technical coefficient as follows:

Producing sector No.	**Total output of the sector**	**Input requirements of producing sectors**				**Requirements for final uses**
		X_1	**X_2**	**X_3**	**X_4**	
1	**2** Sales →	**3**	**4**	**5**	**6**	**7**
1	X_1	$a_{11}x_1$	$a_{12}x_2$	$a_{13}x_3$	$a_{14}x_4$	F_1
2	X_2	$a_{21}x_1$	$a_{22}x_2$	$a_{23}x_3$	$a_{24}x_4$	F_1
3	X_3	$a_{31}x_1$	$a_{32}x_2$	$a_{33}x_3$	$a_{34}x_4$	F_1
4	X_4	$a_{41}x_1$	$a_{42}x_2$	$a_{43}x_3$	$a_{44}x_4$	F_1
Primary input (Labour)	Total primary input = L →	l_1x_1	l_2x_2	l_3x_3	l_4x_4	

It should be noted that all these coefficients are non - negative (> 0)

The above table gives us the total output of each sector in terms of technical coefficients; and if there are *"n"* producing sectors:

$X_1 = a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + \ldots\ldots\ldots\ldots\ldots\ldots\ldots a_{1n}X_n + F_1$

$X_2 = a_{21}X_1 + a_{22}X_2 + a_{23}X_3 + \ldots\ldots\ldots\ldots\ldots\ldots\ldots a_{2n}X_n + F_2$

$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$

$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$

$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$

$X_n = a_{n1}X_1 + a_{n2}X_2 + a_{n3}X_3 + \ldots\ldots\ldots\ldots\ldots\ldots\ldots a_{nn}X_n + F_n$

$L = l_1X_1 + l_2X_2 + l_3X_3 + l_4X_4$

$X_i = \sum_{j=1}^{n} a_{ij}X_j + F_j$; (i = 1, 2, 3,…………..n) and, $L = \sum l_i x_i$

The equations may be put in matrix notations:

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots\cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots\cdots & a_{2n} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & & \cdots\cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots\cdots & a_{nn} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} + \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{pmatrix}$$

X = AX + F and $L = \sum l_i x_i$

or, $X - AX = F$

or $(I-A)X = F$

Now Pre-multiplying both sides of above equation by $(I-A)^{-1}$, we get

$A)^{-1}\ (I-A)X = (I-A)^{-1} F$

or $X = (I-A)^{-1} F$

Thus, given the input coefficient matrix A, the consistent sector output vector X can be estimated for given vector of final demand of sectoral output.

Now let us take a numerical example to show the application of input-output model. Let us consider an economy with three sectors - primary, secondary and tertiary - producing an output of Rs.700 crores, Rs. 800 crores and Rs. 600 crores respectively in a particular year. The flow of output for intermediate use and final consumption is indicated in the following table.

Flow of Output (Rs. Crores)

Sector	Primary	Secondary	Tertiary	Final Demand	Gross Output
Primary	35	200	72	393	700
Secondary	105	96	120	479	800
Tertiary	70	120	90	320	600

The input coefficient a_{ij} 's is derived by taking the ratio of the elements of the column of a particular sector to its output.

$$\text{ie;}\quad A = \begin{bmatrix} \frac{35}{700} & \frac{200}{800} & \frac{72}{600} \\ \frac{105}{700} & \frac{96}{800} & \frac{120}{600} \\ \frac{70}{700} & \frac{120}{800} & \frac{90}{600} \end{bmatrix} = \begin{bmatrix} 0.05 & 0.25 & 0.12 \\ 0.15 & 0.12 & 0.20 \\ 0.10 & 0.15 & 0.15 \end{bmatrix}$$

$$I - A = \begin{bmatrix} (1-0.05) & (0-0.25) & (0-0.12) \\ (0-0.15) & (1-0.12) & (0-0.20) \\ (0-0.10) & (0-0.15) & (1-0.15) \end{bmatrix} = \begin{bmatrix} 0.95 & -0.25 & -0.12 \\ -0.15 & 0.88 & -0.20 \\ -0.10 & -0.15 & 0.85 \end{bmatrix}$$

Since the vector of sectoral output is given by

$X = (I\text{-}A)^{-1} F$,

We have to first find out the inverse of (I - A). Now

$$|(I-A)| = 0.95\ (0.748\text{-}0.03) + 0.25\ (\text{-}0.1275\text{-}0.02) - 0.12\ (0.0225\text{+}0.088)$$

$$= 0.6821 - 0.036875 - 0.01326$$

$$= 0.631965$$

$$\text{Now co-factor of (I-A)} = \begin{bmatrix} 0.718 & 0.1475 & 0.1105 \\ 0.2305 & 0.7955 & 0.1675 \\ 0.1556 & 0.208 & 0.7985 \end{bmatrix}$$

$$\text{Adj (I -A)} = \begin{bmatrix} 0.718 & 0.2305 & 0.1556 \\ 0.1475 & 0.7955 & 0.208 \\ 0.1105 & 0.1676 & 0.7985 \end{bmatrix}$$

$$(I - A)^{-1} = \frac{\text{Adj}\ (I-A)}{|(I-A)|} = \frac{1}{0.631965} \begin{bmatrix} 0.718 & 0.2305 & 0.1556 \\ 0.1475 & 0.7955 & 0.208 \\ 0.1105 & 0.1676 & 0.7985 \end{bmatrix} \begin{bmatrix} 393 \\ 479 \\ 320 \end{bmatrix}$$

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \frac{1}{0.631965} \begin{bmatrix} 28.17 + 110.41 + 49.79 \\ 57.97 + 381.04 + 66.56 \\ 43.43 + 80.23 + 255.52 \end{bmatrix}$$

$$= \frac{1}{0.631965} \begin{bmatrix} 442.37 \\ 505.57 \\ 379.18 \end{bmatrix} = \begin{bmatrix} 700 \\ 800 \\ 600 \end{bmatrix}$$

$X_1 = 700$; $X_2 = 800$ and $X_3 = 600$

These values of X_i's tally with the sectoral output of X_1, X_2 and X_3 shown in the above table.

The same result can be obtained by using Cramer's rule

1.2.4 Dynamic Open Input-Output Analysis

Matrix algebra has another important application in relation to dynamic input-output analysis. The consistent level of sectoral output is given by $X = (I-A)^{-1}F$ where F is the final demand vector consisting of household consumption, government consumption, investment demand, inventory demand, export and import. All these components of final demand are assumed to be autonomous in the static open input-output model. But the investment demand is closely related to the sectoral output. An increase in output calls for additional investment demand. So it is more appropriate to consider the investment demand as endogenous variable instead of exogenous (or autonomous) variable as done in static input-output model. Thus, when we treat the input-output system taking investment demand as endogenous variable, it becomes dynamic input-output model. The mathematical structure of dynamic input-output model is presented below. In addition to the usual notation X_i, X_{ij} and F_i in static input-output model we take an additional symbol I_{ij} which indicates flow of output from sector i to sector j for investment. So the basic equation of the n sector dynamic input output model is given by,

$$X_i = \sum_{j=1}^{n} x_{ij} + \sum_{j=1}^{n} I_{ij} + F_i$$

Where $F_i = HC_i + GC_i + Ch\ S_i + E_i - M_i$ $(i = 1, 2,....n)$

For different n sectors, the sectoral output flow is given by the following system of n equations

$$X_1 = x_{11} + x_{12} + \ldots\ldots\ldots\ldots\ldots + x_{1n} + I_{11} + I_{12} + \ldots\ldots\ldots\ldots\ldots\ldots\ldots.. + I_{1n} + F_1$$

$$X_2 = x_{21} + x_{22} + \ldots\ldots\ldots\ldots\ldots + x_{2n} + I_{21} + I_{22} + \ldots\ldots\ldots\ldots\ldots\ldots\ldots.. + I_{2n} + F_2$$

$$\vdots$$

$$X_n = x_{n1} + x_{n2} + \ldots\ldots\ldots\ldots\ldots + x_{nn} + I_{n1} + I_{n2} + \ldots\ldots\ldots\ldots\ldots\ldots\ldots.. + I_{nn} + F_n$$

Since investment demand is related to the increase in output, we have $I_{ij} = b_{ij}\,\Delta X$, where b_{ij} is capital coefficient indicating amount of output of sector i needed for investment in sector j per unit of j th sector output and ΔX_j represents inverse in output of sector j. Moreover defining the intermediate flow as $x_{ij} = a_{ij} X_j$, we rewrite the above system of equations as

$$X_1 = a_{11}X_1 + a_{12}X_2 + \ldots\ldots\ldots + a_{1n}X_n + b_{11}\,\Delta X_1 + b_{12}\,\Delta X_2 + \ldots\ldots\ldots\ldots\ldots + b_{1n}\,\Delta X_n + F_1$$

$$X_2 = a_{21}X_1 + a_{22}X_2 + \ldots\ldots\ldots\ldots + a_{2n}X_n + b_{21}\,\Delta X_1 + b_{22}\,\Delta X_2 + \ldots\ \ldots\ldots\ldots.. + b_{2n}\,\Delta X_n + F_2$$

$$\vdots$$

$$X_n = a_{n1}X_1 + a_{n2}X_2 + \ldots\ldots\ldots\ldots\ldots + an_nX_n + b_n\,\Delta X_1 + b_{n2}\,\Delta X_2 + \ldots\ldots\ldots\ldots\ldots + b_{nn}\,\Delta X_n + F_n$$

In matrix form, the above set of simultaneous equations can be written as

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots\cdots & a_{2n} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots\cdots & \cdots\cdots \\ a_{n1} & a_{n2} & \cdots\cdots & a_{nn} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ \vdots \\ X_n \end{pmatrix} +$$

$$\begin{pmatrix} b_{11} & b_{12} & \cdots\cdots & b_{1n} \\ b_{21} & b_{22} & \cdots\cdots & b_{2n} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots\cdots & \cdots\cdots \\ b_{n1} & b_{n2} & \cdots\cdots & b_{nn} \end{pmatrix} \begin{pmatrix} \Delta X_1 \\ \Delta X_2 \\ \vdots \\ \vdots \\ \Delta X_n \end{pmatrix} + \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ \vdots \\ F_n \end{pmatrix}$$

or in matrix notation

$X = AX + B \,.\, \Delta X + F$

Now defining $\frac{\Delta X_i}{x_i} = g_i$ as the sectoral growth of output of sector I, we can write

$\Delta X_i = g_i X_i$

In matrix form $\Delta X = GX$,

Where G is the diagonal matrix of sectoral growth rate of output such that

$$\begin{pmatrix} g_1 & 0 & \cdots & 0 \\ 0 & g_2 & \cdots\cdots & 0 \\ \cdots\cdots & \cdots\cdots & \cdots\cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots\cdots & \cdots\cdots \\ 0 & 0 & \cdots\cdots & g_n \end{pmatrix}$$

Now substituting $\Delta X = GX$ in $X = AX + B \,.\, \Delta X + F$, we get

$X = AX + BGX + F$

or $X - AX - BGX = F$

or $(I-A-BG)X = F$

Premultiplying both sides of the above equation by $(I-A-BG)^{-1}$, we have

$(I-A-BG)^{-1} \; (I-A-BG)X = \; (I-A-BG)^{-1} F$

or $X = \; (I-A-BG)^{-1} F$

Thus the above equation gives the consistent level of sectoral output in a dynamic input-output system endogenising the investment demand.

Illustration 1.2.1

Find the consistent output level of a three-sector economy X_1, X_2 and X_3 given the input coefficient matrix (A), capital coefficient matrix (B), diagonal matrix of sectoral growth rate (G) and final demand vector (F).

$$A = \begin{bmatrix} 0.2 & 0.1 & 0.2 \\ 0.3 & 0.3 & 0.2 \\ 0.2 & 0.2 & 0.2 \end{bmatrix}; \qquad B = \begin{bmatrix} 0.1 & 0.2 & 0.1 \\ 0.2 & 0.1 & 0.2 \\ 0.1 & 0.1 & 0.1 \end{bmatrix}$$

$$G = \begin{bmatrix} 0.02 & 0 & 0 \\ 0 & 0.03 & 0 \\ 0 & 0 & 0.02 \end{bmatrix} \qquad F = \begin{bmatrix} 200 \\ 300 \\ 250 \end{bmatrix}$$

Solution : The solution of dynamic input-output model is given by

$X = (I\text{-}A\text{-}BG)^{-1}F$ where

X = Output Vector

I = Identity Matrix

A = Input Coefficient Matrix

B = Capital Coefficient Matrix

G = Diagonal Matrix of growth rate of Sectoral Output

F = Final Demand (excluding Investment Demand)

Substituting the given matrices, we have

$$I\text{-}A\text{-}BG = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.2 & 0.1 & 0.2 \\ 0.3 & 0.3 & 0.2 \\ 0.2 & 0.2 & 0.2 \end{bmatrix} - \begin{bmatrix} 0.1 & 0.2 & 0.1 \\ 0.2 & 0.1 & 0.2 \\ 0.1 & 0.1 & 0.1 \end{bmatrix} \begin{bmatrix} 0.02 & 0 & 0 \\ 0 & 0.03 & 0 \\ 0 & 0 & 0.02 \end{bmatrix}$$

$$= \begin{bmatrix} 0.798 & -0.106 & -0.202 \\ -0.304 & 0.697 & -0.204 \\ -0.202 & -0.203 & 0.798 \end{bmatrix}$$

$|(I - A - BG)|$ = 0.798 (0.556 - 0.041) - 0.106 (0.243 + 0.041) - 0.202 (0.062 + 0.141)

= 0.411 - 0.030 - 0.041

= 0.34

Now

$$\text{Co-factor of } (I - A - BG) = \begin{bmatrix} 0.515 & 0.284 & 0.203 \\ 0.126 & 0.596 & 0.183 \\ 0.162 & 0.224 & 0.524 \end{bmatrix}$$

$$\text{Adj } (I - A - BG) = \begin{bmatrix} 0.515 & 0.126 & 0.162 \\ 0.284 & 0.596 & 0.224 \\ 0.203 & 0.183 & 0.524 \end{bmatrix}$$

$$(I - A - BG)\text{-}1 = \frac{Adj\ (I-A-BG)}{|(I-A-BG)|} = \frac{\begin{bmatrix} 0.515 & 0.126 & 0.162 \\ 0.284 & 0.596 & 0.224 \\ 0.203 & 0.183 & 0.524 \end{bmatrix}}{0.34}$$

$$= \begin{bmatrix} 1.515 & 0.371 & 0.476 \\ 0.835 & 1.753 & 0.659 \\ 0.597 & 0.538 & 1.541 \end{bmatrix}$$

$X = (I\text{-}A\text{-}BG)^{-1} F$

or

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} 1.515 & 0.371 & 0.476 \\ 0.835 & 1.753 & 0.659 \\ 0.597 & 0.538 & 1.541 \end{pmatrix} \begin{pmatrix} 200 \\ 300 \\ 250 \end{pmatrix}$$

or X_1 = 1.515 x 200 + 0.371 x 300 + 0.476 x 250 = 533.30

X_2 = 0.835 x 200 + 1.753 x 300 + 0.659 x 250 = 857.65

X_3 = 0.597 x 200 + 0.538 x 300 + 1.541 x 250 = 666.05

X_1 = 533.30

X_2 = 857.65

X_3 = 666.05

1.2.5 The Hawkins-Simon Condition

Many a time input-output matrix solution may give outputs expressed by negative numbers. If our solution gives negative outputs, it means that more than one ton (or any unit)of that product is used up in the production of every one ton of that product; which is definitely an unrealistic solution. Such a system is not a viable system.

Hawkins-Simon conditions guard against such eventualities.

Our basic equation is $X = (I - A)^{-1} F$

Here, (I−A) must be non-singular (i.e., it must have an inverse), and the inverse must yield non-negative outputs (since negative production makes no sense). This is where the Hawkins-Simon Condition comes in.

The basic equation is $X = (I - A)^{-1} F$, in order that this does not give negative numbers as a solution, the matrix $(I - A)$, which in fact is:

$$\begin{pmatrix} (1-a_{11}) & -a_{12} & -a_{13} & \cdots\cdots & a_{1n} \\ -a_{21} & (1-a_{22}) & -a_{23} & \cdots\cdots & -a_{2n} \\ -a_{31} & -a_{32} & (1-a_{33}) & & -a_{3n} \\ \ldots\ldots & \ldots\ldots. & \ldots\ldots\ldots & & \ldots\ldots. \\ -a_{n1} & -a_{n2} & -a_{n3} & \cdots\cdots & (1-a_{nn}) \end{pmatrix}$$

Should be such that:

The determinant of the matrix must always be positive

The diagonal elements : $(1-a_{11})$, $(1-a_{22})$, $(1-a_{33})$ …………………… $(1-a_{nn})$ should all be positive or, in other words, elements : a_{22}, a_{33} …………………… a_{nn} should all be less than one. Thus one unit of output of any sector should use not more than 1 unit of its own output.

These are called Hawkins-Simon Conditions.

The Hawkins-Simon Condition gives the necessary and sufficient condition for the system $(I-A)X = F$ to have a unique and economically meaningful solution (i.e., a solution with positive output levels for a given positive final demand).

The matrix I−A is non-singular ($|I-A| \neq 0$) and the inverse (I−A)$^{-1}$ exists and is non-negative (i.e., all elements ≥ 0) if and only if all leading principal minors of I−A are positive.

Illustration 1.2.2

Suppose $[A] = \begin{bmatrix} 0.8 & 0.2 \\ 0.9 & 0.7 \end{bmatrix}$

$[I-A] = \begin{bmatrix} 0.2 & -0.2 \\ -0.9 & 0.3 \end{bmatrix}$

and the values of determinant of $[I-A]$ will be 0.06-0.18 = -0.12 which is less than zero.

As such Hawkins-Simon conditions are not satisfied. No solution will be possible in this case.

Illustration 1.2.3

The following inter-industry transactions table was constructed for an economy for the year 1990.

Industry	1	2	Final Consumption	Total
1	500	1600	400	2500
2	1750	1600	4650	8000
Labour	250	4800	-	5050
Total	2500	8000	5050	15550

Construct technology coefficient matrix showing direct requirements. Does a solution exist for this system?

Solution

Technology matrix showing direct requirements per rupee of output is obtained by dividing each input by the total output of the sector.

That is,

$$a_{11} = \frac{X_{11}}{X_1} = \frac{500}{2500} = 0.20$$

$$a_{12} = \frac{X_{12}}{X_2} = \frac{1600}{8000} = 0.20$$

$$a_{21} = \frac{X_{21}}{X_1} = \frac{1750}{2500} = 0.70$$

$$a_{22} = \frac{X_{22}}{X_2} = \frac{1600}{8000} = 0.20$$

Hence technology matrix is given by:

ʌ –

Industry	**1**	**2**
1	0.20	0.20
2	0.70	0.20
Labour	0.10	0.60

and $[I - A] = \begin{bmatrix} 1 - 0.20 & -0.20 \\ -0.70 & 1 - 0.20 \end{bmatrix} = \begin{bmatrix} 0.80 & -0.20 \\ -0.70 & 0.80 \end{bmatrix}$

i.e; $|I - A| = \begin{vmatrix} 0.80 & -0.20 \\ -0.70 & 0.80 \end{vmatrix}$

= 0.80 x 0.80 - 0.20 x 0.70 = 0.50

Since $|I - A|$ is positive and all elements of the principal diagonal of $[I - A]$ are also positive, the Hawkins-Simon conditions are satisfied. Hence, the empirical sytem has a solution.

Illustration 1.2.4

Given :

A= $\begin{bmatrix} 0.1 & 0.3 & 0.1 \\ 0 & 0.2 & 0.2 \\ 0 & 0 & 0.3 \end{bmatrix}$ and final demands are F_1, F_2 and F_3. Find the output levels consistent with the model. What will be the output levels if F_1 = 20, F_2 = 0 and F_3 = 100 ?

Solution:

We know that

$$= \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = (I-A)^{-1} \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix}$$

Now, $[I - A] = \begin{bmatrix} 0.9 & -0.3 & -0.1 \\ 0 & 0.8 & -0.2 \\ 0 & 0 & 0.7 \end{bmatrix}$

Co-factors are follows:

$$A_{11} = + \begin{vmatrix} 0.8 & -0.2 \\ 0 & 0.7 \end{vmatrix} = 0.56$$

$$A_{12} = - \begin{vmatrix} 0 & -0.2 \\ 0 & 0.7 \end{vmatrix} = 0$$

$$A_{13} = + \begin{vmatrix} 0 & 0.8 \\ 0 & 0 \end{vmatrix} = 0$$

$$A_{21} = - \begin{vmatrix} -0.3 & -0.1 \\ 0 & 0.7 \end{vmatrix} = 0.21$$

$$A_{22} = + \begin{vmatrix} 0.9 & -0.1 \\ 0 & 0.7 \end{vmatrix} = 0.63$$

$$A_{23} = - \begin{vmatrix} 0.9 & -0.3 \\ 0 & 0 \end{vmatrix} = 0$$

$$A_{31} = + \begin{vmatrix} -0.3 & -0.1 \\ 0.8 & -0.2 \end{vmatrix} = 0.14$$

$$A_{32} = - \begin{vmatrix} 0.9 & -0.1 \\ 0.8 & -0.2 \end{vmatrix} = 0.18$$

$$A_{33} = + \begin{vmatrix} 0.9 & -0.3 \\ 0 & 0.5 \end{vmatrix} = 0.72$$

Hence the value of the determinant is 0.9 x 0.56 = 0.504

Hence, $(I\text{-}A)^{-1} = \frac{1}{0.504} \begin{bmatrix} 0.56 & 0.21 & 0.14 \\ 0 & 0.63 & 0.18 \\ 0 & 0 & 0.72 \end{bmatrix}$

$$= \begin{bmatrix} 1.11 & 0.42 & 0.28 \\ 0 & 1.25 & 0.36 \\ 0 & 0 & 1.43 \end{bmatrix}$$

i.e; $\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 1.11 & 0.42 & 0.28 \\ 0 & 1.25 & 0.36 \\ 0 & 0 & 1.43 \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix}$

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 1.11F_1 + & 0.42F_2 + & 0.28F_3 \\ 0 + & 1.25F_2 + & 0.36F_3 \\ 0 + & 0 + & 1.43F_3 \end{bmatrix}$$

From the given values of F_1, F_2 and F_3, we have

$X_1 = 1.11F_1 + 0.42F_2 + 0.28F_3$

= 1.11 x 20 + 0 + 0.28 x 100

= 50.2

$X_2 = 1.25F_2 + 0.36F_3$

= 0 + 0.36 x 100

= 36

$X_3 = 1.43F_3$

= 143

It is to be noted that if the technology matrix is upper triangular, i.e; if all elements below the main diagonal are zero or nearly zero, then $(I\text{-}A)^{-1}$ will also be triangular. In such cases, output X_3 depends on only the final demand of sector 3 and X_2 on final demands of sector 2 and 3.

1.2.6 Open and Closed Input - Output Models

In input-output models, an open model considers final demand (consumption, investment, etc.) as exogenous, while a closed model treats all production as intermediate demand within the system, with no external consumption. An open model includes a separate sector for final demand, which represents consumption, investment, government spending, and net exports and a closed model assumes that all output is used as input within the system, meaning there is no final demand.

Our model contains exogenous sector of final demand which supplies primary input factors (labour services - which are not produced by n industries) and consumes the outputs of the n-producing industries (not as input). Such an input-output model is known as open model. It includes exogenous sectors in terms of 'final demand bill' - along with the endogenous sectors in terms of n-producing sectors. Input-output model which has endogenous final demand vector is known as closed input-output model.

Coefficient Matrix and Open Model

Our open model in matrix notations is given by:

$$X = AX + F$$

Where

A = Input Coefficient Matrix

F = Final Demand

X = Total Output Matrix

The input coefficient matrix represented by $[a_{ij}]$

$$= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots\cdots & a_{2n} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ a_{n1} & a_{n2} & \cdots\cdots & a_{nn} \end{pmatrix}$$

is of great importance. Each column of this matrix specifies the input requirements for the production of one unit of a particular commodity. The second column, for example, stated that to produce a unit of commodity 2, the inputs needed are a_{12} units of commodity 1, a_{21}, units of commodity 2, a_{32} units of commodity 3, and a_{n2} units of commodity n.

If no industry uses its own product as an input then,

$$a_{ij} = \begin{pmatrix} 0 & a_{12} & a_{13} & \cdots\cdots & a_{1n} \\ a_{21} & 0 & a_{23} & \cdots\cdots & a_{2n} \\ a_{31} & a_{32} & 0 & \cdots\cdots & a_{3n} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & & \cdots\cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots\cdots & 0 \end{pmatrix}$$

Elements of matrix will be zero whenever the sectors do not trade with each other. It should be noted that no coefficient can be negative, there can be no negative inputs.

Coefficient Matrix in Value Terms

Again, if we assume prices of all the outputs to be given, the ijth elements of $[a_{ij}]$ matrix will represent the amount of ith commodity in money terms needed for producing 'a rupee worth' of jth commodity. For example, if $a_{ij} = 0.35$, it means that 35 paise worth of ith commodity is required as an input producing a rupee worth of jth commodity.

Also, in view of the presence of exogenous sector (which supplies primary inputs) the sum of the elements of each input coefficient column $[a_{ij}]$ must be less than 1. Each column-sum represents the partial input cost (excluding the cost of primary input) incurred in producing a rupee worth of some commodity; if this sum is greater than or equal to one rupee, the production will not be economically justifiable. Symbolically this fact may be stated as:

$\sum_{i=1}^{n} a_{ij} < 1$ (j=1,2,....,n) and each a_{ij} is non-negative, i.e., either zero or greater than zero.

The cost of the primary inputs (which is also termed as value added) needed in producing a unit of jth commodity would be $\left(1 - \sum_{i=1}^{n} a_{ij}\right)$

Note here that a_{ij}'s are in value terms.

Solution of Open Model

Let us consider an economy with n-industries. If producing sector 1 is to produce an output just sufficient to meet the input requirements of the n industries as well as the final demand of the exogenous sector, its output level X_1 must satisfy the following equations:

$$X_1 = a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + \ldots\ldots\ldots\ldots a_{1n}X_n + F_1$$

$$\text{or } (1- a_{11})X_1 - a_{12}X_2 - a_{13}X_3 \ldots\ldots\ldots\ldots (-)\, a_{1n}X_n = F_1$$

For the entire set of n industries, the correct output levels, therefore, can be symbolized by the following set of n linear equations:

$$\begin{array}{ccccccc}
(1-a_{11})X_1 & -a_{12}X_2 & -a_{13}X_3 & - & \cdots\cdots & - & a_{1n}X_n = F_1 \\
-a_{21}X_1 + & (1-a_{22})X_2 & -a_{23}\,X_3 & - & \cdots\cdots & - & -a_{2n}X_n = F_2 \\
\cdots\cdots & \cdots\cdots & \cdots\cdots\cdots & & & & \cdots\cdots\cdots \\
\cdots\cdots & \cdots\cdots & \cdots\cdots\cdots & & & & \cdots\cdots \\
-a_{n1}X_1 & -a_{n2}X_2 & -a_{n3}X_3 & - & \cdots\cdots & + & (1-a_{nn})X_n = F_n
\end{array}$$

In the matrix notation this may be written as:

$$\begin{pmatrix}
(1-a_{11}) & -a_{12} & -a_{13} & \cdots\cdots & a_{1n} \\
-a_{21} & (1-a_{22}) & -a_{23} & \cdots\cdots & -a_{2n} \\
-a_{31} & -a_{32} & (1-a_{33}) & & -a_{3n} \\
\cdots\cdots & \cdots\cdots & \cdots\cdots\cdots & & \cdots\cdots \\
\cdots\cdots & \cdots\cdots & \cdots\cdots\cdots & & \cdots\cdots \\
-a_{n1} & -a_{n2} & -a_{n3}\cdots & & (1-a_{nn})
\end{pmatrix}
\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ \vdots \\ X_n \end{pmatrix}
=
\begin{pmatrix} F_1 \\ F_2 \\ F_3 \\ \vdots \\ \vdots \\ F_n \end{pmatrix}$$

A)X = F

that is, $X = (I-A)^{-1} F$

$X = AX + F$

or, $X - AX = F$

or $(I-A)X = F$

or $X = (I-A)^{-1} F$

Where

A = The given matrix of input coefficients

X = Vectors of output of producing sector

F = Vectors of final demand of producing sector

If $|I - A| \neq 0$, then $(I-A)^{-1}$ exists, we can then estimate for either of the two matrices X and F by assuming one of them to be given exogenously.

It is here that we observe that assumptions made in input-output analysis go a long way in making the problem simplified.

For example, with the assumption of linear homogeneous function, it is possible to write a linear equation of each producing sector which then can be easily transformed into matrix notation.

On the other hand, as long as the input coefficients remain fixed (as assumed) the matrix A will not change or (I-A) will not change. Therefore, in finding the solution of X = (I-A)$^{-1}$ only one matrix inversion needs to be performed even if we are to consider thousands of different final demand vectors according to alternative development targets. Hence, such an assumption of fixed technical coefficients has meant considerable saving in computational effort.

Illustration 1.2.5

Suppose there are only three industries in an economy and we have to estimate the output of each (sector) industry with the given input coefficient matrix and final demand as follows (the coefficient matrix is in value terms):

$$A = \begin{array}{c} \begin{array}{ccc} P & Q & R \end{array} \\ \begin{bmatrix} 0.3 & 0.4 & 0.2 \\ 0.2 & 0 & 0.5 \\ 0.1 & 0.3 & 0.1 \end{bmatrix} \end{array} \text{ and } F= \begin{bmatrix} 100 \\ 40 \\ 50 \end{bmatrix} \text{million rupees}$$

Here we note that 3 column sums of A are (0.3 + 0.2 + 0.1) = 0.6, (0.4 + 0 + 0.3) = 0.7 and (0.2 + 0.5 + 0.1) = 0.8; which are less than 1 in each case. In other words, (1-0.6) = 0.4, (1 - 0.7) = 0.3 and (1 - 0.8) = 0.2 is the maximum amount of primary input which can be used for producing 'a rupee worth' of the three commodities (P, Q and R) respectively.

$$\text{Since } A = \begin{bmatrix} 0.3 & 0.4 & 0.2 \\ 0.2 & 0 & 0.5 \\ 0.1 & 0.3 & 0.1 \end{bmatrix}; [I - A] = \begin{bmatrix} +0.7 & -0.4 & -0.2 \\ -0.2 & +1 & -0.5 \\ -0.1 & -0.3 & +0.9 \end{bmatrix}$$

Substituting these values in X = $[I - A]^{-1}$ F, we get:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} +0.7 & -0.4 & -0.2 \\ -0.2 & +1 & -0.5 \\ -0.1 & -0.3 & +0.9 \end{bmatrix} \begin{bmatrix} 100 \\ 40 \\ 50 \end{bmatrix}$$

$$\text{But} \begin{bmatrix} +0.7 & -0.4 & -0.2 \\ -0.2 & +1 & -0.5 \\ -0.1 & -0.3 & +0.9 \end{bmatrix} = \frac{1}{0.401} \begin{bmatrix} 0.75 & 0.42 & 0.22 \\ 0.23 & 0.61 & 0.39 \\ 0.16 & 0.25 & 0.62 \end{bmatrix}$$

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \frac{1}{0.401} \begin{bmatrix} 0.75 & 0.42 & 0.22 \\ 0.23 & 0.61 & 0.39 \\ 0.16 & 0.25 & 0.62 \end{bmatrix} \begin{bmatrix} 100 \\ 40 \\ 50 \end{bmatrix}$$

$X_1 = \frac{1}{0.401}$ { 0.75 (100) + 0.42 (40) + 0.22 (50)}

= Rs. 279 million (approx.)

$X_2 = \frac{1}{0.401}$ { 0.23 (100) + 0.61 (40) + 0.39 (50)}

= Rs. 167 million (approx.), and

$X_3 = \frac{1}{0.401}\{0.16\,(100) + 0.25\,(40) + 0.62\,(50)\}$

= Rs. 142 million (approx.)

Coefficient Matrix and Closed Model

We shall now examine whether we will be able to estimate F or X if the model is changed into closed one. If the exogenous sector (final demand bill) of the open input-output model is absorbed into the system of endogenous sectors, the model would turn into a closed one. In such a model final demand bill and primary inputs will not appear any more; rather in their place, we shall have the input requirements and output of this newly conceived industry, the 'household industry' producing the primary input labour. Final demand sector would now be considered as one of endogenous sectors. As such now we shall have (n + 1) industries in place of n industries and all producing for the sake of satisfying the input requirements.

This newly conceived industry (of final demand bill) will also be assumed to have a fixed input ratio as any other industry. In other words, the supply of primary input must now bear a fixed proportion to final demand (i.e., consumption of this newly conceived industry). This will mean, for example, that households will consume each commodity in fixed proportion to the labour services they supply.

Looking at the problem in this particular way, it appears that the conversion of open model into a closed one should not create any significant change in our analysis and solution because disappearance of final demand means only an addition of one more homogeneous equation to already existing set of n homogeneous equations. Is it really so? Let us examine.

Let us assume that there are four industries only - including the new one (of final demand) designated by subscript 0. We shall, therefore, have the following set of equations:

$X_0 = a_{00}X_0 + a_{01}X_1 + a_{02}X_2 + a_{03}X_3$

$X_1 = a_{10}X_0 + a_{11}X_1 + a_{12}X_2 + a_{13}X_3$

$X_2 = a_{20}X_0 + a_{21}X_1 + a_{22}X_2 + a_{23}X_3$

$X_3 = a_{30}X_0 + a_{31}X_1 + a_{32}X_2 + a_{33}X_3$

This gives us a homogeneous equation system

$$\begin{pmatrix} (1-a_{11}) & -a_{12} & -a_{13} & \cdots\cdots & a_{1n} \\ -a_{21} & (1-a_{22}) & -a_{23} & \cdots\cdots & -a_{2n} \\ -a_{31} & -a_{32} & (1-a_{33}) & & -a_{3n} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & & \cdots\cdots \\ -a_{n1} & -a_{n2} & -a_{n3}\cdots & & (1-a_{nn}) \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} F_1 \\ F_2 \\ F_3 \\ \vdots \\ \vdots \\ F_n \end{pmatrix}$$

$[\boldsymbol{I} - \boldsymbol{A}]\,\boldsymbol{X} = \mathbf{0}$

Since the 4 rows of the input coefficient matrix happen to be linearly dependent, $|I - A|$ will turn out to be zero. Hence, the solution is indeterminate.

This means that in a closed model no unique output-mix of each sector exists. We can at most determine the output levels of endogenous sectors in proportion to one another, but cannot fix their absolute levels unless additional information is made available exogenously.

We reach the same conclusion by analyzing the model in a slightly different way.

Let us now assume a competitive economy of three producing sectors. We transform the output and inputs of each sector into its receipts and payments by taking into consideration the prices which are endogenously determined:

Receipts: S.I $P_1 X_1 = P_1 a_{11}X_1 + P_1 a_{12}X_2 + P_1 a_{13}X_3 + P_1F_1$ .... (1)

S.II $P_2 X_2 = P_2 a_{21}X_1 + P_2 a_{22}X_2 + P_2 a_{23}X_3 + P_2F_2$ .... (2)

S.III $P_3 X_3 = P_3a_{31}X_1 + P_3a_{32}X_2 + P_3a_{33}X_3 + P_3F_3$ .... (3)

Payments: S.I $P_1 X_1 = P_1 a_{11}X_1 + P_2 X_{21}X_1 + P_3 a_{31}X_1 + WL_1 + rK_1$ .... (4)

S.II $P_2 X_2 = P_1 a_{12}X_2 + P_2 X_{22}X_2 + P_3 a_{32}X_2 + WL_2 + rK_2$ .... (5)

S.III $P_3 X_3 = P_1 a_{13}X_3 + P_2 X_{23}X_3 + P_3 a_{33}X_3 + WL_3 + rK_3$ .... (6)

Where W is the wage rate and r is the capital return; L_1, L_2, K_1, K_2 and K_3 stand for labour and capital employed in each sector.

Therefore, $(P_1 a_{11}X_1 + P_2 a_{21}X_1 + P_3 a_{31}X_1)$, $(P_1 a_{12}X_2 + P_2 a_{22}X_2 + P_3 a_{32}X_3)$ and $(P_1 a_{13}X_3 + P_2 a_{23}X_3 + P_3 a_{33}X_3)$ are the payments to intermediate inputs and $(WL_1 + rK_1)$, $(WL_2 + rK_2)$ and $(WL_3 + rK_3)$ are the payments to capital and labour in each sector.

Also NNP = $Y = P_1F_1 + P_2F_2 + P_3F_3 = W(L_1 + L_2 + L_3) + r(K_1 + K_2 + K_3)$ (This also represents the value added)

Thus, in the whole system we have only 7 equations while these involve 18 variables: (X_1, X_2, X_3), (F_1, F_2, F_3) (K_1, K_2, K_3), (P_1, P_2, P_3) (W, r) and Y. Therefore, the system is indeterminate. Now, if we transform the model into a closed one by assuming that labour and capital are used in fixed proportions to output:

i.e; $L_1 = U_1 X_1$

$L_2 = U_2 X_2$

$L_3 = U_3 X_3$

and

$K_1 = V_1 X_1$

$K_2 = V_2 X_2$

$K_3 = V_3 X_3$

Still the number of equations is less than the variables involved in the closed model. Hence the closed model does not give us the required solution until we have some variables exogenously given.

Example : 1

Let us consider an economy with three production sectors and households. The transactions (inter-industry and household consumption) are as follows (in Rs. crores):

	Sector 1	**Sector 2**	**Sector 3**	**Households**	**Total Output**
Sector 1	30	20	10	40	100
Sector 2	25	15	20	40	100
Sector 3	15	30	25	30	100
Households	30	30	30	10	100

Here we need to estimate the output of each sector and the households consume output from 1,2 and 3. The households also supply labour to all sectors and there is no final demand from outside, i.e; it is a closed system. The total output of each sector is:

$$A = \begin{bmatrix} 0.30 & 0.20 & 0.10 & 0.40 \\ 0.25 & 0.15 & 0.20 & 0.40 \\ 0.15 & 0.30 & 0.25 & 0.30 \\ 0.30 & 0.30 & 0.30 & 0.10 \end{bmatrix}$$

In a closed model, the equilibrium condition is

$$AX = X = (I - A)X = 0$$

But since this gives only homogeneous equations, we add a normalization condition, e.g.:

$x_1 + x_2 + x_3 + x_4 = 400$ (Assumed Total Output)

$$I - A = \begin{bmatrix} 0.70 & -0.20 & -0.10 & -0.40 \\ -0.25 & 0.85 & -0.20 & -0.40 \\ -0.15 & -0.30 & 0.75 & -0.30 \\ -0.30 & -0.30 & -0.30 & 0.90 \end{bmatrix}$$

Let this matrix be M, then solve:

$MX = 0$ with $x_1 + x_2 + x_3 + x_4 = 400$

Expected Outcome $X = \begin{bmatrix} 95.0 \\ 105.0 \\ 100.0 \\ 100.0 \end{bmatrix}$

These values would balance all inter-industry and household transactions.

x_1 = Rs. 95 cores : Sector 1 must produce Rs. 95 crore worth of output

x_2 = Rs.105 crores: Sector 2 must produce Rs.105 crores

x_3 = Rs. 100 crores: Sector 3 must produce Rs. 100 crores

x_4 = Rs. 100 crores: Households contribute Rs. 100 crores in labour, wages and consumption

This closed model ensures self-contained equilibrium, where households are both consumers and providers of input (labour/income), and all production is used up within the system.

Limitations of Input-Output Analysis

1. Errors in forecasting final demand will have grave consequences.
2. Current relative prices of inputs may not be same as the ones implied in the table.
3. The assumption of linear homogeneous production function may not be valid. The technical coefficients will not remain constant even if input price ratios are held constant in such circumstances.
4. The constant coefficient formulation ignores the possibility of industry output reaching capacity, changing prices and input proportions in the table.
5. The assumption of constant technical coefficients goes counter to the possibility of substitution of inputs.
6. Sectoral division is, for practical purposes, limited. Such a sectorisation is not good enough for many forecasting purposes.
7. Sectorisation (grouping of commodities in sectors) is often arbitrary. The intra-sectoral heterogeneity with respect to technologies, efficiency and demand is not invariant over time.
8. Regional input-output analysis involves many more assumptions and difficulties in construction of such tables.

Summarised Overview

Matrices serve as a fundamental tool in economics, particularly in analysing the interdependence among sectors of an economy through input-output models. These models, introduced by Wassily W Leontief, provide a structured, quantitative framework to understand how output from one industry becomes input for another, thus capturing the circular flow of economic activity. The static input-output model represents economic relationships at a given point in time, while the dynamic model includes time-based changes such as technological advancement or capital accumulation. Open models treat final demand (like household consumption) as exogenous, while closed models endogenies sectors such as households within the system. The Hawkins-Simon condition is a mathematical criterion used to check the feasibility and viability of input-output systems. It ensures that each sector can produce non-negative outputs to meet the demand, given the inter-sectoral dependencies. In contemporary economic planning, matrices and input-output models have found wide-ranging applications from policy-making, regional planning, and environmental analysis to employment estimation, supply chain optimisation, and impact assessment of economic shocks or digital transformations.

Assignments

1. Explain the structure and significance of an input-output model in economic analysis. How does matrix algebra facilitate this analysis? Illustrate with a simple numerical example.

2. Differentiate between static and dynamic input-output models. Discuss their assumptions, applications, and limitations with appropriate illustrations.

3. Compare and contrast open and closed input-output models. How does the inclusion of the household sector in the closed model impact the overall system? Support your answer with a diagram or table.

4. Define the Hawkins-Simon condition. Why is this condition important for the viability of an input-output model? Illustrate with a 2×2 matrix and interpret your findings.

5. Using a hypothetical input-output table of three sectors (Agriculture, Manufacturing, and Services), construct the technical coefficients matrix and calculate the gross output required to meet a given final demand.

References

1. Yamane, Taro. (2012). *Mathematics for Economists: An Elementary Survey*. New Delhi:Prentice Hall of India.

2. Sreenath Baruah(2012): *Basic Mathematics and its applications in Economics*, Macmillan India Ltd.

3. Mehta and Madnani (2022): *Mathematics for Economists*, Sultan Chand & Sons, New Delhi

Suggested Readings

1. Chiang, A.C. (2008), *Fundamental Methods of Mathematical Economics*, McGraw Hill, New York.

2. Y.P. Agarwal: *Statistical Methods: Concepts, Application and Computation*, Sterling Publishers 1986

3. Hooda R.P: *Statistics for Business and Economics* , Mac Million, New Delhi

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

3

UNIT

Determinants and Hessian Matrix

Learning Outcomes

After completing this unit, the learner will be able to :

- understand determinants and their properties
- compute determinants of higher-order matrices
- apply Hessian determinants in optimisation

Background

Have you ever tried solving a puzzle where the pieces fit together in a very specific way, one wrong move, and everything falls apart? Mathematics often works like that too. When you are dealing with systems of equations or complex problems in economics or physics, every number has its place, and every method has its purpose. One such powerful tool that helps in organising and solving such puzzles is the idea of matrices and their special friend, the determinant.

Imagine you are trying to solve a mystery where every clue is hidden in a table of numbers. Some tables lead you to the answer directly, while others are trickier and need deeper thinking. Determinants help you understand whether the clues (or equations) are leading somewhere meaningful, or going in circles. It is like checking if a maze has a way out before you even enter it.

Now, as we move further into the world of mathematics, the problems get a little more interesting. What happens when there are more variables, more connections, and more complexity? This is where learning to handle bigger tables of numbers, called higher-order matrices, becomes important. Evaluating their determinants is not just about applying a formula, it is about understanding the structure and balance hidden within them.

And sometimes, you do not just want to solve equations, you want to know if you are reaching the best solution. That is where another powerful tool steps in, the Hessian matrix. It helps us test whether we have truly found a maximum or minimum in a problem, like double-checking if the treasure we have found is really the biggest prize.

Keywords

Matrix, Determinant, Laplace Method, Hessian Determinant

1.3.1 Higher Order Determinants

Determinants are mathematical tools used to analyse square matrices, providing valuable information about their properties. Higher-order determinants, which extend beyond 2×2 or 3×3 matrices, can be evaluated using the Laplace expansion method, which involves breaking down a determinant into smaller components along any row or column. The Hessian determinant, a specific application in calculus, represents the determinant of the Hessian matrix - a matrix of second-order partial derivatives. It plays a crucial role in optimisation problems, helping identify the nature of critical points in multivariable functions, such as whether they represent maxima, minima, or saddle points.

The determinant of a 3×3 matrix $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ is called a third-order determinant and is the summation of three products.

Take the first element a_{11} and delete the remaining elements in the first row and first column. Multiply a_{11} by the remaining determinant $\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$.

Take the second element a_{12} and delete the remaining elements in the first row and second column. Multiply a_{12} by the remaining determinant $\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$.

Take the third element a_{13} and delete the remaining elements in the first row and third column. Multiply a_{13} by the remaining determinant $\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$.

Thus, $|A| = a_{11} \times \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$

$$= a_{11} \times (a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

In like manner, the determinant of a 4×4 matrix is the sum of four products, the determinant of a 5×5 matrix is the sum of five products etc.

1.3.2 Properties of Determinants

Property 1

The value of the determinant remains unchanged, if the determinant's rows are changed into columns and the columns into rows. That is, $|A| = |A'|$

For example,

Let the determinant be, $\begin{vmatrix} 5 & 4 \\ 1 & 2 \end{vmatrix}$. Value of the determinant,

$$\begin{vmatrix} 5 & 4 \\ 1 & 2 \end{vmatrix} = (5 \times 2) - (4 \times 1) = 10 - 4 = 6$$

If the rows and columns of the determinant are interchanged, the initial determinant becomes, $\begin{vmatrix} 5 & 1 \\ 4 & 2 \end{vmatrix}$

The value of the new determinant is, $\begin{vmatrix} 5 & 1 \\ 4 & 2 \end{vmatrix} = (5 \times 2) - (4 \times 1) = 10 - 4 = 6$

Property 2

Sign of the determinant changes when any two rows or columns are interchanged.

Let the determinant be $\begin{vmatrix} 2 & 5 & 9 \\ 3 & 1 & 2 \\ 7 & 4 & 6 \end{vmatrix}$. The value of the given determinant is,

$$\begin{vmatrix} 2 & 5 & 9 \\ 3 & 1 & 2 \\ 7 & 4 & 6 \end{vmatrix} = 2\,(1 \times 6 - 4 \times 2) - 5\,(3 \times 6 - 7 \times 2) + 9\,(3 \times 4 - 7 \times 1)$$

$$= 2\,(6 - 8) - 5\,(18 - 14) + 9\,(12 - 7)$$

$$= 2\,(-2) - 5\,(4) + 9\,(5)$$

$$= -4 - 20 + 45$$

$$= 21$$

If we interchange the first row with the second row $(R_1 \rightarrow R_3)$, the new determinant will be, $\begin{vmatrix} 7 & 4 & 6 \\ 3 & 1 & 2 \\ 2 & 5 & 9 \end{vmatrix}$

Let us find the value of the new determinant.

$$\begin{vmatrix} 7 & 4 & 6 \\ 3 & 1 & 2 \\ 2 & 5 & 9 \end{vmatrix} = 7\,(1 \times 9 - 5 \times 2) - 4\,(3 \times 9 - 2 \times 2) + 6\,(3 \times 5 - 1 \times 2)$$

$$= 7\,(9 - 10) - 4\,(27 - 4) + 6\,(15 - 2)$$

$$= 7\,(-1) - 4\,(23) + 6\,(13)$$

$= -7 - 92 + 78$

$= -21$

Thus, the sign of the determinant changes when the two rows or columns are interchanged.

Property 3

The value of the determinant is zero if all elements of a row or column are zero.

For example,

Let the determinant be $\begin{vmatrix} 7 & 4 & 6 \\ 0 & 0 & 0 \\ 2 & 5 & 9 \end{vmatrix}$. Then the value of the determinant will be,

$$\begin{vmatrix} 3 & 4 & 6 \\ 0 & 0 & 0 \\ 2 & 5 & 2 \end{vmatrix} = 3\,(0 \times 2 - 0 \times 5) - 4\,(0 \times 2 - 0 \times 2) + 6\,(0 \times 5 - 0 \times 2)$$

$$= 3\,(0) - 4\,(0) + 6\,(0) = 0$$

Property 4

If any two rows or columns are identical the value of the determinant is zero.

For example,

Let the determinant be $\begin{vmatrix} 1 & 3 & 1 \\ 4 & 2 & 5 \\ 1 & 3 & 1 \end{vmatrix}$. Then the value of the determinant will be,

$$\begin{vmatrix} 1 & 3 & 1 \\ 4 & 2 & 5 \\ 1 & 3 & 1 \end{vmatrix} = 1\,(2 \times 1 - 3 \times 5) - 3\,(4 \times 1 - 1 \times 5) + 1\,(4 \times 3 - 1 \times 2)$$

$$= 1\,(2 - 15) - 3\,(4 - 5) + 1\,(12 - 2)$$

$$= 1\,(-13) - 3\,(-1) + 1(10) = -13 + 3 + 10 = 0$$

Property 5

If all the elements of a row or column of a determinant are multiplied by a constant k the value of the determinant gets multiplied by k.

For example,

Let the given determinant be $\begin{vmatrix} 2 & 1 & 4 \\ 3 & 1 & 5 \\ 2 & 4 & 1 \end{vmatrix}$. Then the value of the determinant will be,

$$\begin{vmatrix} 2 & 1 & 4 \\ 3 & 1 & 5 \\ 2 & 4 & 1 \end{vmatrix} = 2\,(1 \times 1 - 4 \times 5) - 1\,(3 \times 1 - 2 \times 5) + 4\,(3 \times 4 - 1 \times 2)$$

$$= 2\,(1 - 20) - 1\,(3 - 10) + 4\,(12 - 2)$$

$= 2(-19) - 1(-7) + 4(10)$

$= -38 + 7 + 40 = 9$

Let us multiply the first row of the given determinant by a constant 2. $That\ is\ R_1 \times 2.$

$$\begin{vmatrix} 2x2 & 2x1 & 2x4 \\ 3 & 1 & 5 \\ 2 & 4 & 1 \end{vmatrix} = \begin{vmatrix} 4 & 2 & 8 \\ 3 & 1 & 5 \\ 2 & 4 & 1 \end{vmatrix}$$

The value of the new determinant will be,

$$\begin{vmatrix} 4 & 2 & 8 \\ 3 & 1 & 5 \\ 2 & 4 & 1 \end{vmatrix} = 4(1 \times 1 - 4 \times 5) - 2(3 \times 1 - 2 \times 5) + 8(3 \times 4 - 1 \times 2)$$

$$= 4(1 - 20) - 2(3 - 10) + 8(12 - 2)$$

$$= 4(-19) - 2(-7) + 8(10)$$

$$= -76 + 14 + 80$$

$$= 18$$

The value of the initial determinant was 9. Multiplying the first row initial determinant with a constant 2 leads to a new determinant value 18 which is a constant times the value of the initial determinant. That is $2x9 = 18.$

Property 6

For a triangular determinant, the value of the determinant will be the product of the leading diagonals.

For example,

The value of the given determinant $\begin{vmatrix} 4 & 0 & 0 \\ 3 & 1 & 0 \\ 7 & 2 & 9 \end{vmatrix}$ will be,

$$\begin{vmatrix} 4 & 0 & 0 \\ 3 & 1 & 0 \\ 7 & 2 & 9 \end{vmatrix} = 4(1 \times 9 - 2 \times 0) - 0(3 \times 9 - 7 \times 0) + 0(3 \times 2 - 1 \times 7)$$

$$= 4(9 - 0) - 0(27 - 0) + 0(6 - 7)$$

$$= 4(9) - 0(27) + 0(-1)$$

$$= 36$$

Let us find the product of the leading diagonal elements now.

$$\begin{vmatrix} 4 & 0 & 0 \\ 3 & 1 & 0 \\ 7 & 2 & 9 \end{vmatrix} \rightarrow 4 \times 1 \times 9 = 36$$

Thus, for a triangular determinant, the value of the determinant will be the product of the leading diagonals.

1.3.3 Laplace Expansion and Higher – Order Determinants

Laplace expansion is a method for evaluating determinants in terms of cofactors. It thus simplifies matters by permitting higher-order determinants to be established in terms of lower-order determinants. Laplace expansion of a third-order determinant can be expressed as

$$|A| = a_{11}|C_{11}| + a_{12}|C_{11}| + a_{13}|C_{13}|$$

where C_{ij} is a cofactor based on a second-order determinant.

Illustration 1.3.1

Use Laplace Expansion to find the determinant for the matrix A expanding along the second column

$$\begin{bmatrix} 15 & 7 & 9 \\ 2 & 5 & 6 \\ 9 & 0 & 12 \end{bmatrix}$$

Solution

$$\text{Let } A = \begin{bmatrix} 15 & 7 & 9 \\ 2 & 5 & 6 \\ 9 & 0 & 12 \end{bmatrix}$$

$$|A| = a_{12}|C_{12}| + a_{22}|C_{22}| + a_{32}|C_{32}|$$

$$a_{12} = 7, \quad a_{22} = 5, \qquad a_{32} = 0$$

$$C_{12} = -\begin{vmatrix} 2 & 6 \\ 9 & 12 \end{vmatrix} = -(24 - 54) = 30$$

$$C_{22} = +\begin{vmatrix} 15 & 9 \\ 9 & 12 \end{vmatrix} = (180 - 81) = 99$$

$$C_{32} = +\begin{vmatrix} 15 & 9 \\ 2 & 6 \end{vmatrix} = -(90 - 18) = -72$$

$$|A| = 7 \times 30 + 5 \times 99 + 0 \times -72$$

$$= 210 + 495$$

$$= 705$$

Illustration 1.3.2

Use Laplace Expansion to find the determinant for the matrix A expanding along the second row.

$$\begin{bmatrix} 2 & 4 & 1 & 5 \\ 3 & 2 & 5 & 1 \\ 1 & 2 & 1 & 4 \\ 3 & 4 & 3 & 2 \end{bmatrix}$$

Solution

$$A = \begin{bmatrix} 2 & 4 & 1 & 5 \\ 3 & 2 & 5 & 1 \\ 1 & 2 & 1 & 4 \\ 3 & 4 & 3 & 2 \end{bmatrix}$$

$$|A| = a_{21}|C_{21}| + a_{22}|C_{22}| + a_{23}|C_{23}| + a_{24}|C_{24}|$$

$$a_{21} = 3, \; a_{22} = 2, a_{23} = 5, \; a_{24} = 1$$

$$C_{21} = -3 \times \begin{vmatrix} 4 & 1 & 5 \\ 2 & 1 & 4 \\ 4 & 3 & 2 \end{vmatrix} = -3[4(2-12) - 1(4-16) + 5(6-4)]$$

$$= -3[4 \times -10 - 1 \times -12 + 5 \times 2$$

$$= -3[-40 + 12 + 10]$$

$$= -3 \times -18$$

$$= 54$$

$$C_{22} = 2 \times \begin{vmatrix} 2 & 1 & 5 \\ 1 & 1 & 4 \\ 3 & 3 & 2 \end{vmatrix} = 2[2(2-12) - 1(2-12) + 5(3-3)]$$

$$= 2[2 \times -10 - 1 \times -10 + 5 \times 0$$

$$= 2[-20 + 10]$$

$$= 2 \times -10$$

$$= -20$$

$$C_{23} = 5 \times \begin{vmatrix} 2 & 4 & 5 \\ 1 & 2 & 4 \\ 3 & 4 & 2 \end{vmatrix} = -5[2(4-16) - 4(2-12) + 5(4-6)]$$

$$= -5[2 \times -12 - 4 \times -10 + 5 \times -2$$

$$= -5[-24 + 40 - 10]$$

$$= -5 \times 6$$

$$= -30$$

$$C_{24} = 1 \times \begin{vmatrix} 2 & 4 & 1 \\ 1 & 2 & 1 \\ 3 & 4 & 3 \end{vmatrix} = [2(6-4) - 4(3-3) + 1(4-6)]$$

$$= [2 \times 2 - 4 \times 0 + 1 \times -2$$

$$= [4 - 0 - 2]$$

$$= 2$$

$$|A| = 54 - 20 - 30 + 1 \times 2$$

$$= 6$$

Illustration 1.3.3

Use Laplace Expansion to find the determinant for the matrix A expanding along the first column.

$$\begin{bmatrix} 5 & 0 & 1 & 3 \\ 4 & 2 & 6 & 0 \\ 3 & 0 & 1 & 5 \\ 0 & 1 & 4 & 2 \end{bmatrix}$$

Solution

$$A = \begin{bmatrix} 5 & 0 & 1 & 3 \\ 4 & 2 & 6 & 0 \\ 3 & 0 & 1 & 5 \\ 0 & 1 & 4 & 2 \end{bmatrix}$$

$$|A| = a_{11}|C_{11}| + a_{21}|C_{21}| + a_{31}|C_{31}| + a_{41}|C_{41}|$$

$$a_{21} = 5,\ \ a_{22} = 4, a_{23} = 3,\ a_{24} = 0$$

$$C_{11} = 5 \times \begin{vmatrix} 2 & 6 & 0 \\ 0 & 1 & 5 \\ 1 & 4 & 2 \end{vmatrix} = 5[2(2-20) - 6(0-5) + 0(0-1)]$$

$$= 5[2 \times -18 - 6 \times -5 + 0$$

$$= 5[-36 + 30]$$

$$= -30$$

$$C_{21} = 4 \times \begin{vmatrix} 0 & 1 & 3 \\ 0 & 1 & 5 \\ 1 & 4 & 2 \end{vmatrix} = -4[0 \times (2-20) - 1(0-5) + 3(0-1)]$$

$$= -4[-1 \times -5 + 3 \times -1$$

$$= -4[5 - 3]$$

$$= -4 \times 2$$

$$= -8$$

$$C_{31} = 3 \times \begin{vmatrix} 0 & 1 & 3 \\ 2 & 6 & 0 \\ 1 & 4 & 2 \end{vmatrix} = 3[0(12 - 0) - 1(4 - 0) + 3(8 - 6)]$$

$$= 3[0 - 1 \times 4 + 3 \times 2$$

$$= 3[-4 + 6]$$

$$= 3 \times 2$$

$$= 6$$

$$C_{41} = 0 \times \begin{vmatrix} 0 & 1 & 3 \\ 2 & 6 & 0 \\ 0 & 1 & 5 \end{vmatrix} = 0$$

$$|A| = -30 - 8 + 6$$

= -32

1.3.4 Hessian Matrix

The Hessian matrix is a square matrix of second-order partial derivatives of a scalar-valued function. It is widely used in optimisation and numerical analysis to study the local curvature of a function.

It is widely used in economics to analyse the curvature of functions, particularly in optimisation problems. It helps determine whether a function has a local maximum, local minimum, or saddle point by examining the determinant of the Hessian. In consumer theory, it is applied to assess utility maximisation or expenditure minimisation, while in production theory, it evaluates profit maximisation or cost minimisation.

Given $z = f(x, y)$. The first order condition for maxima or minima (optimisation) is

$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} = 0$, i.e. $z_x = z_y = 0$

The sufficient condition for the function $z = f(x, y)$ is maximum and minimum is

1. $z_{xx} > 0$ for minimum, $z_{xx} < 0$for maximum

2. $z_{xx} . z_{yy} > (z_{xy})^2$, where $z_{xx} = \frac{\partial^2 z}{\partial x^2}$, $z_{yy} = \frac{\partial^2 z}{\partial y^2}$, $z_{xy} = \frac{\partial^2 z}{\partial x \partial y}$

A convenient test for this second-order condition is the Hessian.

The Hessian determinant $|H| = \begin{vmatrix} z_{xx} & z_{xy} \\ z_{yx} & z_{yy} \end{vmatrix}$

If the first element on the principal diagonal, the first principal $|H_1| = z_{xx} > 0$, and $|H_2| = |H| > 0$, Hessian is called positive definite. A positive definite Hessian satisfies the second-order conditions for a minimum.

If the first principal minor $|H_1| = z_{xx} < 0$, $|H_2| = |H| > 0$ Hessian is called negative definite. A negative definite Hessian satisfies the second-order conditions for a maximum.

Illustration 1.3.4

Show that the function $z = 3x^2 - xy + 2y^2 - 4x - 7y + 12$ is optimised using the Hessian to test the second order condition.

Solution

$$z = 3x^2 - xy + 2y^2 - 4x - 7y + 12$$

$$\frac{\partial z}{\partial x} = 6x - y - 4, \quad \frac{\partial z}{\partial y} = -x + 4y - 7$$

$$\frac{\partial^2 z}{\partial x^2} = 6, \quad \frac{\partial^2 z}{\partial y^2} = 4, \quad \frac{\partial^2 z}{\partial x \partial y} = -1$$

$$|H_1| > 0, \quad |H_2| = \begin{vmatrix} z_{xx} & z_{xy} \\ z_{yx} & z_{yy} \end{vmatrix}$$

$$= \begin{vmatrix} 6 & -1 \\ -1 & 4 \end{vmatrix} = 24 - 1 = 23 > 0$$

$$|H_1| > 0, \quad |H_2| > 0 \text{ and } \frac{\partial^2 z}{\partial x^2} = 6 > 0$$

$|H|$ is positive definite. z is minimised at the critical values.

Illustration 1.3.5

Show that the function $z = 2x^2 + 5xy + 8y^2$ is optimised using the Hessian to test the second order condition.

Solution

$$z = 2x^2 + 5xy + 8y^2$$

$$\frac{\partial z}{\partial x} = 4x + 5y, \quad \frac{\partial z}{\partial y} = 5x + 16y$$

$$\frac{\partial^2 z}{\partial x^2} = 4, \quad \frac{\partial^2 z}{\partial y^2} = 16, \quad \frac{\partial^2 z}{\partial x \partial y} = 5$$

$$|H_1| > 0, \quad |H_2| = \begin{vmatrix} z_{xx} & z_{xy} \\ z_{yx} & z_{yy} \end{vmatrix}$$

$$= \begin{vmatrix} 4 & 5 \\ 5 & 16 \end{vmatrix} = 64 - 25 = 39 > 0$$

$|H_1| > 0, \quad |H_2| > 0$ and $\frac{\partial^2 z}{\partial x^2} = 4 > 0$

$|H|$ is positive definite. z is minimised at the critical values.

Illustration 1.3.6

A firm produces two goods in pure competition and has the following total revenue and total cost functions.

$$TR = 15Q_1 + 18Q_2, \qquad TC = 2Q_1^2 + 2Q_1Q_2 + 3Q_2^2$$

The two goods are technically related in production. Since the marginal cost of one is dependent on the level of output of the other. Maximise profits for the firm using Hessian for the second-order condition.

Solution

$$\pi = TR - TC = 15Q_1 + 18Q_2 - 2Q_1^2 - 2Q_1Q_2 - 3Q_2^2$$

$$\frac{\partial \pi}{\partial Q_1} = 15 - 4Q_1 - 2Q_2, \quad \frac{\partial \pi}{\partial Q_2} = 18 - 2Q_1 - 6Q_2$$

$$\frac{\partial^2 \pi}{\partial Q_1^2} = -4, \quad \frac{\partial^2 \pi}{\partial Q_2^2} = -6, \quad \frac{\partial^2 \pi}{\partial Q_1 \partial Q_2} = -2$$

$$|H_1| = -4 < 0,$$

$$|H_2| = \begin{vmatrix} -4 & -2 \\ -2 & -6 \end{vmatrix} = 24 - 4 = 20 > 0$$

$|H_1| < 0, \quad |H_2| > 0$ and $\frac{\partial^2 \pi}{\partial Q_1^2} = -4 < 0$

$|H|$ is negative definite. z is maximised at the critical values.

Summarised Overview

Determinants are essential tools in mathematics used to analyse square matrices, helping in the study of their properties. They provide valuable insights into the matrix's structure, such as whether a system of linear equations has a unique solution or not. Higher-order determinants, which involve matrices larger than 2x2 or 3x3, are typically computed using methods like Laplace expansion, where the determinant is broken down into smaller components along any row or column. This technique helps in evaluating complex determinants by simplifying them into manageable parts. Additionally, the Hessian matrix plays a significant role in optimisation problems, particularly in determining the nature of critical points in multivariable functions. The Hessian is used to identify whether a function has a local maximum, minimum, or saddle point by analysing its second-order partial derivatives.

Key properties of determinants include their invariance when rows are swapped with columns, and the sign change that occurs when two rows or columns are interchanged. The determinant is also zero if any row or column consists entirely of zeros or if two rows or columns are identical. When rows or columns are multiplied by a constant, the determinant's value changes by the same constant factor. In triangular matrices, the determinant equals the product of the diagonal elements. The Laplace expansion method simplifies the process of calculating higher-order determinants by expanding along any row or column and utilising cofactors. Lastly, the Hessian matrix, formed by the second-order partial derivatives of a function, is crucial in optimisation for testing second-order conditions, helping determine if a critical point is a minimum or maximum. This matrix is applied across various fields, including economics, to optimise functions such as utility maximisation and cost minimisation.

Assignments

1. Use Laplace Expansion to find the determinant for the matrix A expanding along the first column.

$$\begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 2 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 2 & 1 & 0 \end{bmatrix}$$

2. Use Laplace Expansion to find the determinant for the matrix A expanding along the first column.

$$\begin{bmatrix} 3 & 2 & 5 & 7 \\ -1 & -4 & -3 & 0 \\ 6 & 4 & 2 & -1 \\ 2 & -1 & 0 & 3 \end{bmatrix}$$

3. Show that the function $z = -3x^2 + 4xy - 4y^2$ is optimised using the Hessian to test the second order condition.

References

1. Anton, H., & Rorres, C. (2014). *Elementary Linear Algebra* (11th ed.). Wiley.
2. Ghosh, P. (2015). *Mathematical Methods for Economists* (2nd ed.). Oxford University Press.
3. Rao, S. S. (2017). *Optimization Theory and Applications* (4th ed.). Wiley.
4. Bhattacharyya, G. K., & Johnson, R. A. (2012). *Statistical Concepts and Methods* (2nd ed.). Wiley.

Suggested Readings

1. Yamane, Taro. (2012). *Mathematics for Economists: An Elementary Survey*. New Delhi: Prentice Hall of India.
2. Chiang, A.C. (2008), *Fundamental Methods of Mathematical Economics*, McGraw Hill, New York.
3. Marsden, J. E., & Tromba, A. J. (2003). *Vector Calculus* (5th ed.). W. H. Freeman.
4. Tuckwell, H. C. (2011). *Optimization and Control* (1st ed.). Springer.
5. Chandler, D., & Taylor, G. (2009). *Introduction to Optimization* (3rd ed.). Pearson Education.
6. Sundaram, R. K. (2010). *A First Course in Optimization Theory* (1st ed.). Cambridge University Press.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

2

BLOCK

Differential and Difference Equations

1

UNIT

First Order Differential Equations

Learning Outcomes

After completing this unit, the learner will able to:

- understand the concept and classification of differential equations
- identify and solve first order differential equations
- apply the first order differential equations to economic scenarios
- differentiate between limited and unlimited growth

Background

In economics, mathematical tools and techniques play a crucial role in simplifying dynamic relationships and providing analytical clarity to economic behavior over time. Among these tools, differential equations hold a special place, especially when modeling economic processes that involve continuous change. These equations enable economists to describe how one variable changes in relation to another, typically time which is essential for understanding and forecasting economic phenomena. First-order differential equations, the simplest type of differential equations, are particularly useful in economic analysis. They model a variety of dynamic situations where the rate of change of a variable is dependent on its current state or another influencing variable. These include phenomena like capital accumulation, population growth, inflation dynamics, price adjustments, and resource depletion.

Keywords

First Differential Equation, Exponential Growth, Logistic Growth, Linear Differential Equation

Discussion

2.1.1 First Order Differential Equations - Definitions and Concepts

A first-order differential equation is a type of differential equation that involves only the first derivative of the unknown function with respect to an independent variable. It typically represents the rate of change of quantity and is widely used in modelling various real-world processes where changes over time or space are studied. Mathematically, it is expressed in the form $\frac{dy}{dx} = f(x, y)$, where y is the dependent variable and x is the independent variable. Since it contains only the first derivative, it is called a first-order equation. These equations are fundamental in understanding dynamic systems in economics, providing a basis for modeling growth, decay, equilibrium, and adjustment processes. Depending on their structure, first-order differential equations can be solved using methods such as separation of variables, integrating factors, or substitution techniques.

2.1.1.1 Differential Equations

A differential equation is an equation that involves one or more derivatives of a dependent variable with respect to one or more independent variables. It is an implicit functional relationship between variables and their differentials. The order of a differential equation is the order of the highest derivative in the equation. A differential equation is of order n if the n^{th} derivative is the highest order derivative in it. The degree of a differential equation is the degree of the highest derivative in the equation. A differential equation is an equation which expresses an explicit or implicit relationship between a function $y = f\ (t)$ and one or more of its derivatives or differentials. Examples of differential equations include

$$\frac{dy}{dt} = 7t + 12$$

$$y' = \frac{dy}{dt} = 14y$$

$$y'' - 2y' + 21 = 0$$

Equations involving a single independent variable, such as those above, are called ordinary differential equations. The solution or integral of a differential equation is any equation, without derivative or differential, that is defined over an interval and satisfies the differential equation for all the values of the independent variable(s) in the interval. Ordinary differential equations involve only one independent variable and its derivative. When there are several independent variables we may have partial differential equations. To solve a differential equation means to find all its solutions and in case it has no solution, to show it has no solution. A solution to a differential equation is any function which satisfies the equation i.e., reduces it to an identity when substituted into the equation. A differential equation of order n has a general solution containing n arbitrary independent constants. A particular solution may be obtained from the general solution by assigning certain specific values to the arbitrary constants which are obtained from any given information of the initial conditions given in the problem.

Example.1 : To solve the differential equation $y''(t) = 9$ for all the functions y(t) which satisfy the equation, simply integrate both sides of the equation to find the integrals.

$$y'(t) = \int 9\,dt = 9t + c_1$$

$$y(t) = \int (9\,t + c_1)dt = \frac{9}{2}\,t^2 + c_1 t + c$$

This is called a general solution which indicates that when c is unspecified, a differential equation has an infinite number of possible solutions. If c can be specified, the differential equation has a particular or definite solution which alone of all possible solutions is relevant.

Example.2 :

$\frac{dy}{dt} = 3x + 9$ is a differential equation of order 1, degree 1

$\left(\frac{d^2y}{dt^2}\right)^3 - 9\left(\frac{dy}{dt}\right)^2 + xy$ is a differential equation of order 2, degree 3

$\left(\frac{dy}{dt}\right)^4 - 6t^6 = 0$ is a differential equation of order 1, degree 4

$\frac{d^2y}{dt^2} + \left(\frac{dy}{dt}\right)^3 + x^2 = 0$ is a differential equation of order 2, degree 1

$\left(\frac{d^2y}{dt^2}\right)^8 - 9\left(\frac{d^3y}{dt^3}\right)^5 = 85y$ is a differential equation of order 3, degree 5

Example.3 : To solve the differential equation of the type

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

We can resort here to integration because, if $f\,(x)$ is the derivative of $F(x)$, we can write the above equation

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

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

$y = F(x) + C$, This is the required solution

Example.4 : $\frac{dy}{dx} = y$ may be written $\frac{dy}{y}$ = dx

Integrating $\quad lny = x + c$

or y = e^{x+c} or y = $e^{x} . e^{c}$

If we write $C_1 = e^{c}$ then y = C_1 e^{x} is the solution.

Example.5 : $\frac{dy}{dx}$ = xe^{x} is equivalent to dy = xe^{x} dx, i.e;

y = $\int xe^{x}\, dx = e^{x}$ (x - 1) + C, this is the general solution

If the initial condition is that $y = 2$ when $x = 0$ then by substitution

2 = e^{0} (0-1) + C or 2 = -1 + C

C = 3

The particular solution is y = e^{x} (x - 1) + 3

2.1.2 General formula for Differential Equations

For a first-order linear differential equation, $\frac{dy}{dt}$ and y must be of the first degree, and no product $y\left(\frac{dy}{dt}\right)$ may occur. For such an equation

$$\frac{dy}{dt} + vy = z$$

where v and z may be constants or functions of time, the formula for a general solution is

$$y(t)e^{\int vdt} = \int z\, e^{\int vdt}\,\mathrm{dt} + \mathrm{A}$$

$$y(t) = e^{-\int vdt}\left(\int z\, e^{\int vdt}\,\mathrm{dt} + \mathrm{A}\right)$$

where A is an arbitrary constant. A solution is composed of two parts: $e^{-\int vdt}$A is called the complementary function, and $e^{-\int vdt} \int ze^{\int vdt}$dt is called the particular integral. The particular integral y_p equals the intertemporal equilibrium level of y(t); the complementary function y_c represents the deviation from the equilibrium. For $y(t)$ to be dynamically stable, y_c must approach zero as t approaches infinity (that is, k in e^{kt} must be negative). The solution of a differential equation can always be checked by differentiation.

Illustration 2.1.1

Find the general solution for the differential equation $\frac{dy}{dt} + 4y = 12$

Solution:

Since $v = 4$ and $z = 12$;

$\int v\, dt = \int 4\, dt = 4t + C$

C is always ignored and subsumed under A.

Thus,

y(t) = e^{-4t} ($\int 12e^{4t}$dt+A)

Integrating the remaining integral gives $\int 12e^{4t}$dt = $\frac{12e^{4t}}{4} = 3e^{4t}$ + c

Ignoring the constant again and substituting above

y(t) = e^{-4t} ($3e^{4t} + A$) = Ae^{-4t} + 3

Since $e^{-4t}\ e^{4t} = e^0$ = 1. As t → ∞

y_c = Ae^{-4t} → 0 and y(t) approaches y_p = 3, the intertemporal equilibrium level. y(t) is dynamically stable.

To check this answer, which is a general solution because A has not been specified, start by taking the derivative of y(t) = Ae^{-4t} + 3

$\frac{dy}{dt} = -4Ae^{-4t}$

From the original problem,

$\frac{dy}{dt}$ + 4y =12

$\frac{dy}{dt} = 12 - 4y$

Substituting y = Ae^{-4t} + 3, $\frac{dy}{dt} = 12 - 4\,(Ae^{-4t} + 3) = -4Ae^{-4t}$

Illustration 2.1.2

(a) Solve the equation below using the formula for a general solution.
(b) Check your answer. $2\,\frac{dy}{dt} - 2t^2y = 9t^2 \quad y(0) = -2.5$

Solution:

Dividing through by 2

$\frac{dy}{dt}$ - $t^2y = 4.5t^2$

Thus, $v = -t^2, z = 4.5t^2$,

and $\int v\, dt = \int -t^2 dt$ = -$\frac{1}{3}$$t^3$, substituting,

$y(t) = e^{-\int vdt}\left(\int z\, e^{\int vdt}\ \text{dt} + \text{A}\right)$

$y(t) = e^{-\int -t^2 dt}\left(\int 4.5\, t^2\, e^{\int -t^2 dt}\, \mathrm{dt} + \mathrm{A}\right)$

y(t) = $e^{(1/3)t^3}$ $\int 4.5t^2e^{-(1/3)t^3}$dt + A)….(1)

Let u = $-\left(\frac{1}{3}\right)t^3, \frac{du}{dt} = -t^2$ and $dt = -\frac{du}{t^2}$

Thus

$\int 4.5t^2e^{-(1/3)t^3}$dt = $\int 4.5t^2e^u\frac{du}{-t^2}$ = -4.5 $\int e^u$ du =$-4.5e^{-(1/3)t^3}$

Substituting in (1),

y(t) = $e^{(1/3)t^3}$(- $4.5e^{-(1/3)t^3} + A$) = A $e^{(1/3)t^3}$ - 4.5

Apply $y(0) = -2.5,$ At t = 0, -2.5 = A - 4.5; A = 2. Thus,

y(t) = 2 $e^{(1/3)t^3}$ - 4.5

Taking the derivative of y(t) = 2 $e^{(1/3)t^3}$ - 4.5, we get $\frac{dy}{dt} = 2\,t^2e^{(1/3)t^3}$.

Substituting $\frac{dy}{dt}$ and y in the given equation

2 $\frac{dy}{dt}$ - $2t^2y = 9t^2$, we get

$2\left(2\,t^2e^{\left(\frac{1}{3}\right)t^3}\right) - 2t^2\left(2\,e^{\left(\frac{1}{3}\right)t^3} - 4.5\right) = 9t^2$

At $t = 0, y(0) = 2e^0 - 4.5 = 2 - 4.5 = -2.5$

Exact differential equations and partial integration

Given a function of more than one independent variable, such as F(y, t) where $M = \frac{\partial F}{\partial y}$ and $N = \frac{\partial F}{\partial t}$, the total differential is written as

$$dF(y,t) = M\,dy + N\,dt$$

Since F is a function of more than one independent variable, M and N are partial derivatives and in the above equation is called a partial differential equation. If the differential is set equal to zero, so that M dy + N dt = 0 it is called an exact differential equation because the left side exactly equals the differential of the primitive function F(y, t). For an exact differential equation, $\frac{\partial M}{\partial t}$ must equal $\frac{\partial N}{\partial y}$, that is, $\frac{\partial^2 F}{\partial y \partial t}$.

Solution of an exact differential equation calls for successive integration with respect to one independent variable at a time while holding constant the other independent variable(s). The procedure, called partial integration, reverses the process of partial differentiation.

Illustration 2.1.3

Solve the exact nonlinear differential equation

$$(9y^2t + 12y^3)\,dy + (3y^3 + 8t)\,dt = 0$$

Solution:

Test to see if it is an exact differential equation.

Here M = $9y^2t + 12y^3$ and N = $3y^3 + 8t$.

Thus, ∂M/ ∂t = $9y^2$. ∂N/ ∂y = $9y^2$. If ∂M/ ∂t = ∂N/ ∂y, it is an exact differential equation

Since M = ∂F/ ∂y is a partial derivative, integrate M partially with respect to y by treating t as a constant, and add a new function Z(t) for any additive terms of t which would have been eliminated by the original differentiation with respect to y. Note that ∂y replaces dy in partial integration.

$F(y, t) = \int(9y^2t + 12y^3)\,\partial y + Z(t) = 3y^3t + 3y^4 + Z(t)$ …(1)

This gives the original function except for the unknown additive terms of t, Z(t).

$N = \frac{\partial F}{\partial t}$ is the partial derivative of F with respect to t, treating y as a constant.

Differentiate the above equation with respect to t to find ∂F/ ∂t (earlier called N).

Thus,

$$\frac{\partial F}{\partial t} = 3y^3 + Z'(t)$$

Since $\frac{\partial F}{\partial t}$ = N and N = $3y^3 + 8t$, substitute $\frac{\partial F}{\partial t} = 3y^3 + 8t$ in the above equation

$$3y^3 + 8t = 3y^3 + Z'(t) \qquad Z'(t) = 8t$$

Next integrate Z'(t) with respect to t to find the missing t terms

$$Z(t) = \int Z'(t)\,dt = \int 8t\,dt = 4t^2$$

Substitute the above equation in (1) and add a constant of integration

$F(y, t) = 3y^3t + 3y^4 + 4t^2 + c$

This is easily checked by differentiation

Illustration 2.1.4

Solve the exact nonlinear differential equation

$(24y + 7t + 6)\,dy + (7y + 4t - 9)\,dt = 0$

Solution

Test to see if it is an exact differential equation.

Here $M = 24y + 7t + 6$ and $N = 7y + 4t - 9$. Thus, $\partial M/ \partial t = 7 = \partial N/ \partial y$

$F(y, t) = \int(24y + 7t + 6)\, \partial y + Z(t) = 12y^2 + 7yt + 6y + Z(t)$…(1)

Differentiate the above equation with respect to t to find $\partial F/ \partial t$ (earlier called N). Thus,

$$\frac{\partial F}{\partial t} = 7y + Z'(t), \text{ but } \frac{\partial F}{\partial t} = N = 7y + 4t - 9,$$

so $7y + Z'(t) = 7y + 4t - 9$ $\qquad\qquad Z'(t) = 4t - 9$

4. Next integrate Z’(t) with respect to t to find the missing t terms

$$Z(t) = \int Z'(t)\, dt = \int(4t - 9)\, dt = 2t^2 - 9t$$

5. Substitute the above equation in (1) and add a constant of integration

$F(y, t) = 12y^2 + 7yt + 6y + 2t^2 - 9t + c$

2.1.2.1 Integrating Factors

Not all differential equations are exact. However, some can be made exact by means of an integrating factor. This is a multiplier which permits the equation to be integrated.

Rules for the Integrating Factor

Two rules will help to find the integrating factor for a nonlinear first-order differential equation, if such a factor exists. Assuming $\partial M/ \partial t \neq \partial N/ \partial y$

Rule 1. If $\frac{1}{N}\left(\frac{\partial M}{\partial t} - \frac{\partial N}{\partial y}\right) = functionof\ y\ alone = f(y)$ then $e^{\int f(y)dy}$ is an integrating factor

Rule 2. If $\frac{1}{N}\left(\frac{\partial N}{\partial y} - \frac{\partial M}{\partial t}\right) = functionof\ t\ alone = g(t)$, then $e^{\int g(t)dt}$ is an integrating factor.

Illustration 2.1.5

Testing the nonlinear differential equation $7yt\, dy + (7y^2 + 10t)\, dt = 0$ reveals that it is not exact.

Solution:

With $M = 7yt$ and $N = 7y^2 + 10t$, $\frac{\partial M}{\partial t} = 7y \neq \frac{\partial N}{\partial y} = 14y$. Multiplying by an integrating factor of t, however makes it exact: $7yt^2\, dy + (7y^2 t + 10t^2)\, dt = 0$.

To illustrate the rules above, find the integrating factor, where

$M = 7yt, \quad N = 7y^2 + 10t$

$\frac{\partial M}{\partial t} = 7y \neq \frac{\partial N}{\partial y} = 14y$

Applying Rule 1,

$\frac{1}{7y^2+10t}$ (7y - 14y) = $\frac{-7y}{7y^2+10t}$

which is not a function of y alone and will not supply an integrating factor for the equation. Applying Rule 2,

$\frac{1}{7yt}$ (14y - 7y) = $\frac{7y}{7yt} = \frac{1}{t}$

which is a function of t alone. The integrating factor, therefore, is $e^{\int(1/t)dt} = e^{lnt}$ = t.

Illustration 2.1.6

Use the integrating factors provided in parentheses to solve the following differential equation, 7t dy + 14y dt = 0

Solution:

$M = 7t, \qquad N = 14y$

$\frac{\partial M}{\partial t} = 7 \neq \frac{\partial N}{\partial y} = 14$. The equation is not exact.

But multiplying by the integrating factor t,

$7t^2$ dy + 14yt dt = 0

$\frac{\partial M}{\partial t} = 14t = \frac{\partial N}{\partial y}$, The equation become exact equation.

continuing with the new function

F(y, t) = $\int 7t^2\ \partial y + Z\ (t) = 7t^2y + Z\ (t)$

$\frac{\partial F}{\partial t}$ = 14ty + $Z'\ (t)$ = N = 14ty, so $Z'\ (t) = 0$

Z(t) = $\int 0\ dt$= k, which will be subsumed under the c below

F(y, t) = $7t^2\ y\ +\ c$

2.1.2.2 Separation of Variables

Solution of nonlinear first-order first-degree differential equations is complex. A first-order, first-degree differential equation is one where the highest derivative is $\frac{dy}{dt}$ and it appears to the first power only (i.e., not squared or under a root). It is nonlinear if it contains a product of y and $\frac{dy}{dt}$, or y raised to a power other than 1.) If the equation is exact or can be rendered exact by an integrating factor, the procedure outlined in Example 8 can be used. If, however, the equation can be written in the form of separated variables such that $R(y)\ dy\ +\ S(t)dt\ =\ 0$, where R and S, respectively, are functions of y and t alone, the equation can be solved simply by ordinary integration.

2.1.3 Economic Applications

Differential equations have a wide range of applications in economics. They are used to determine the conditions for dynamic stability in microeconomic models of market equilibrium and to trace the time path of economic growth under various macroeconomic conditions.

Given the growth rate of a variable, differential equations help economists determine the original function. For example, they can be used to estimate demand functions from point elasticity. Similarly, differential equations are applied to derive capital functions from investment models, and to calculate total cost and total revenue from marginal cost and marginal revenue functions.

Illustration 2.1.7

Find the demand function Q = f (P) if point elasticity ε is -1 for all P > 0

Solution:

$\varepsilon = \frac{dQ}{dP}\frac{P}{Q} = -1 \qquad \frac{dQ}{dP} = -\frac{Q}{P}$

Separating the variables

$\frac{dQ}{Q} + \frac{dP}{P} = 0$

Integrating the above gives us:

$\ln Q + \ln P = \ln c$

$QP = c$

$Q = \frac{c}{P}$

Illustration 2.1.8

Find the demand function $Q = f(P)$ if $\varepsilon = -\frac{5P + 2P2}{Q}$ and Q =600 when P = 10

Solution:

$$\varepsilon = \frac{dQ}{dP}\frac{P}{Q} = \frac{-(5P + 2P^2)}{Q}$$

$$\frac{dQ}{dP} = \frac{-(5P + 2P^2)}{Q}\frac{Q}{P} = -(5 + 2P)$$

Separating the variables

$$dQ + (5 + 2P)\, dp = 0$$

Integrating,

$$Q + 5P + P^2 = c$$

$$Q = -P^2 - 5P + c$$

At P =10 and Q = 600,

$600 = -100 - 50 + c, c = 750$

Thus, $Q = 750 - 5P - P^2$

Illustration 2.1.9

The rate of investment is given by I(t) = $140t^{3/4}$ and the initial stock of capital at t = 0 is 250. Determining the function for capital K, the time path K(t).

Solution:

K = $\int 140t^{\frac{3}{4}}$dt = $140 \int t^{\frac{3}{4}}$dt

By the Power rule,

K = 140 $(\frac{4}{7}t^{\frac{7}{4}})$ + c = $80t^{\frac{7}{4}}$ + c

But c = K_0 = 250. Therefore, K = $80t^{\frac{7}{4}}$ + 250.

Illustration 2.1.10

Marginal cost is given by MC = dTC /dQ = 35 + 60Q - $12Q^2$ Fixed cost is 65. Find the (a) Total Cost (b) Average Cost and (c) Variable cost functions.

Solution:

TC = $\int MC$ dQ = $\int(35 + 60Q - 12Q^2)$ dQ = 35Q + $30Q^2$ - $4Q^3$ + c

With FC = 65, at Q = 0, TC = FC = 65. Thus c = FC =65 and

TC = 35Q + $30Q^2$ - $4Q^3$ + c + 65

$$AC = \frac{TC}{Q} = \frac{35Q+30Q^2-4Q^3+65}{Q} = 35 + 30Q - 4Q^2 + \frac{65}{Q}$$

VC = TC - FC = 35Q + $30Q^2$ - $4Q^3$

Illustration 2.1.1 1

Find (a) Total revenue function and (b) the demand function, given

MR = 94 - 4Q - Q^2

TR = $\int MR$ dQ = $\int(94 - 4Q - Q^2)$ dQ = 94Q + $2Q^2$ - $\frac{1}{3}Q^3$ + c

At Q = 0, TR = 0, Therefore c = 0. Thus, TR = 94Q - $2Q^2$ - $\frac{1}{3}Q^3$

P =AR = $\frac{TR}{Q}$ = 94 - 2Q- $\frac{1}{3}Q^2$

2.1.4 Differential Equations for Limited and Unlimited Growth

One of the most important applications of differential equations is in studying the growth of populations, economic output, or resource usage. The two fundamental types of growth patterns observed in real-life situations are: (1) Unlimited growth (2) Limited growth. These growth types differ based on the presence or absence of constraints like resources, space, or external influences.

2.1.4.1 Unlimited rowth

Unlimited growth assumes that the rate of growth of a quantity is directly proportional to its current size, and that there are no restrictions on the resources available for growth. This leads to exponential increases in the quantity over time. It is often used to describe short-term behavior where limitations do not yet have a noticeable impact.

If a quantity or population y grows at a rate proportional that quantity's size, it can be modeled with unlimited growth, which has the differential equation:

$y' = ry$, where r is a constant

This model assumes that the rate of change of a quantity (e.g., population) is directly proportional to the size of the quantity at any time.

$$\frac{dN}{dt} = rN$$

Where,

N(t) → Quantity at time t (eg. Population)

r → Growth rate, $r > 0$ (constant)

$$\frac{dN}{dt} = rdt = \ln |N| = rt + C$$

Exponentiating both sides,

$$N(t) = N_0 e^{rt}$$

where:

- N(t) represents the population or quantity at time t,
- N_0 is the initial population or quantity at t=0,
- r is the constant rate of growth, and
- t is the time variable.

Here, growth is unbounded: As $t \to \infty$, $N(t) \to \infty$ and Applicable when resources are infinite and no limiting factors exist.

Illustration 2.1.1 2

Suppose a bacterial population doubles every 3 hours. If the initial population is 1000, what will be the population after 9 hours?

Since the population doubles every 3 hours, we can use the formula for exponential growth and solve for the growth rate r.

$2 = e^{r.3}$

Taking the natural logarithm of both sides: ln(2)=r·3 r = $\frac{ln(2)}{3}$ = 0.231

$N(t) = N_0e^{rt}$

Substituting the values N_0=1000 , r = 0.231, and t = 9

$= 1000.e^{0.231.9} = 1000.e^{2.079} = 1000.8 = 8000$

Illustration 2.1.1 3

A population grows by 9% each year. If the current population is 6,000, find an equation for the population after t years.

Solution:

The population is growing by a percent of the current population, so this is unlimited growth.

$\frac{dy}{dt} = 0.09y$ Separate the variables

$\frac{1}{y}$ dy = 0.09dt Integrate both sides

$\ln|y| = 0.09t + C$ Exponentiate both sides

$e^{ln|y|} = e^{0.09t + C}$

$y = N(t) = N_0e^{rt}$

$y = 6000\ e^{0.09t}$

Notice that the solution to the unlimited growth equation is an exponential equation

2.1.4.2 Limited growth

In contrast, limited growth considers natural constraints such as food, space, competition, or disease, which limit the growth of a population or system as it becomes larger. This type of growth starts off exponentially, but gradually slows down and eventually levels off at a maximum sustainable size known as the carrying capacity. When a product is advertised heavily, sales will tend to grow very quickly, but eventually the market will reach saturation, and sales will slow. In this type of growth, called limited growth, the population grows at a rate proportional to the distance from the maximum value.

If a quantity grows at a rate proportional to the distance from the maximum value, M, it can be modelled with limited growth, which has the differential equation:

$y' = k(M - y)$

Where,

k is a constant

M is the maximum size of y.

When growth is subject to environmental constraints (e.g., food, space, disease), the population slows down as it approaches a maximum sustainable size called carrying capacity.

$\frac{dN}{dt} = rN\ (1- \frac{N}{K})$

Where,

K is Carrying capacity

r is the intrinsic growth rate

N is the population at time t

This is a non-linear differential equation. Its solution is:

$$N(t) = \frac{K}{1+ (\frac{K-N_0}{N_0})\ e^{-rt}}$$

Here, the growth is initially exponential, slows down as resources deplete and approaches the carrying capacity K as t → ∞

Illustration 2.1.1 4

A population grows logistically with intrinsic growth rate of 0.5, carrying capacity is 1000 and initial population is 100, find population after 5 units of time.

$r\ = 0.5$

$K = 1000$

$N_0 = 100$

The population after 5 units of time:

$$N(5) = \frac{1000}{1+ (\frac{1000-100}{100})\ e^{-0.05.5}} = \frac{1000}{1+9e^{-2.5}} = \frac{1000}{1+9.0.0821} = \frac{1000}{1.739} = 575$$

Illustration 2.1.1 5

A new television is introduced. The company estimates they will sell 200 thousand televisions. After 1 month they have sold 20 thousand. How many will they have sold after 9 months?

Solution:

In this case there is a maximum amount of televisions they expect to sell, so M = 200 thousand. Modeling the sales, y, in thousands of televisions, we can write the differential equation

$y' = k(200 - y)$

Since it was a new television, $y(0) = 0$.We also know the sales after one month,

$y(1) = 20.$

Solving the differential equation,

$\frac{dy}{dx} = k\,(200 - y)$ Separate the variables

$\frac{dy}{200-y} = kdt$ Integrate both sides. On the left use the substitution u = 200-y

$-\ln|200 - y| = kt + C$ Multiply both sides by -1, and exponentiate both sides

$e^{\ln|200-y|} = e^{-kt - C}$ simplify

$y = Be^{-kt}$ Substract 200, divide by -1, and simplify

$y = Ae^{-kt} + 200$

Using the initial condition y(0) = 0

$0 = Ae^{-k(0)} + 200$, So $0 = A + 200$, giving $A = -200$

Using the value y(1) = 20,

$20 = -200e^{-k(1)} + 200$ Substract 200 and divide by -200

$\frac{-180}{-200} = 0.9 = e^{-k}$ Take the ln of both sides

$\ln 0.9 = \ln e^{-k} = -k$ Divide by -1

As a quick sanity check, this value is positive as we would expect, indicating that the sales are growing over time. We now have the equation for the sales of televisions over time:

$$A = -200e^{-0.105t} + 200$$

Finally, we can evaluate this at t = 9 to find the sales after 9 months

$A = -200e^{-0.105t} + 200 = 122.26$ thousand televisions

Limited growth is also commonly used for learning models, since when learning a new skill, people typically learn quickly at first, then their rate of improvement slows down as they approach mastery. Earlier we used unlimited growth to model a population, but often a population will be constrained by food, space, and other resources. When a population grows both proportional to its size, and relative to the distance from some maximum, that is called logistic growth. This leads to the differential equation $y' = ky(M - y)$, which is accurate but not always convenient to use. We will use a slight modification. If a quantity grows at a rate proportional to its size and to the distance from the maximum value, M, it can be modeled with logistic growth, which has the differential equation.

$$y' = ry(1 - \frac{y}{M})$$

r can be interpreted as "the growth rate absent constraints" - how the population would grow if there was not a maximum value.

This differential equation has solutions of the form

$$y(t) = \frac{M}{1 + Ae^{-rt}}$$

Summarised Overview

First-order differential equations form a foundational tool in the analysis of dynamic economic systems. Core techniques such as solving linear and separable equations are applied to key models including exponential and logistic growth. Exponential functions capture scenarios of unrestricted expansion, while logistic models incorporate natural or economic constraints, such as limited resources or saturation points. The price adjustment mechanism illustrates how prices converge toward equilibrium in response to excess demand or supply. Applications extend to modelling population dynamics, capital accumulation, and market adjustments, providing students with robust analytical frameworks to evaluate and forecast economic outcomes.

Assignments

1. Define a first-order differential equation. Give an example.
2. What is the method of separation of variables? Solve $\frac{dy}{dx} = 2xy$
3. Differentiate between exponential and logistic growth models.
4. What is the integrating factor used for solving linear differential equations?
5. Explain the difference between a general solution and a particular solution of a differential equation.

6. How does the exponential growth model represent unlimited growth? What is its mathematical form?

7. Derive the solution of the logistic growth model $\frac{dN}{dt}$ = rN (1- $\frac{N}{K}$) and discuss its economic relevance.

8. A firm's investment grows at a rate proportional to its current value. Formulate the differential equation and solve it for I(0)=500, r=0.05.

9. Compare and contrast the assumptions behind limited and unlimited growth models. Which model is more realistic in the context of developing economies?

10. Suppose demand and supply functions are D(p)=100−2p, S(p) = 20 + 3p. Write and solve the price adjustment differential equation $\frac{dp}{dt}$ = α (D - S)

11. (a)Derive the general solution for the differential equation $\frac{dy}{dt}$ = ky. Interpret the constants involved in the context of an economic variable like capital or population.

    (b) Solve the following differential equation using the integrating factor method:

    $\frac{dy}{dt}$ + 2y =10.

    Illustrate its application in modelling a firm's revenue over time.

    (c) Consider a simple unlimited growth model of national income: $\frac{dy}{dt}$ = rY. If Y(0)=1000 and r=0.05, find Y(t). Plot the growth for t=0 to t=10 and interpret the results.

References

1. Edward T Dowling, Schaum's Outline Series (2001), *Introduction to Mathematical Economics*, Third Edition, McGRAW-HILL

2. Yamane, Taro. (2012). *Mathematics for Economists: An Elementary Survey*. New Delhi: Prentice Hall of India.

3. Sreenath Baruah(2012): *Basic Mathematics and its applications in Economics*, Macmillan India Ltd.

4. Mehta and Madnani (2022): *Mathematics for Economists*, Sultan Chand & Sons, New Delhi

5. Monga G S (2014), *Mathematics and Statistics for Economics*, Vikas Publishing House Pvt Ltd

Suggested Readings

1. Chiang, A.C. (2008), *Fundamental Methods of Mathematical Economics*, McGraw Hill, New York.
2. Y.P. Agarwal: *Statistical Methods: Concepts, Application and Computation*, Sterling Publishers 1986
3. Hooda R.P: *Statistics for Business and Economics*, Mac Million, New Delhi

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

2

UNIT

Solutions of First-Order Linear Difference Equations

Learning Outcomes

After completing this unit, the learner will be able to:

- understand and classify first-order linear difference equations
- solve first-order linear difference equations using the general formula

Background

Ravi, a young entrepreneur, had recently started selling eco-friendly notebooks. In his first month, he sold 100 books. Encouraged by the response, he invested in better marketing. In the second month, he sold 120. By the third, 145. Every month, his sales grew - but not randomly. He noticed a pattern and wondered: Can I predict next month's sales without guessing?

One evening, over a cup of tea, Ravi spoke to his cousin Meera, an economics student. She smiled and said, "You're not just running a business, Ravi. You're watching a mathematical pattern unfold - one that we study in economics to understand how things change over time."

She explained how economists and business analysts use something called difference equations to understand changes that happen in steps - like months, quarters, or years. Unlike changes that happen continuously, these equations look at how things move from one point to the next, just like Ravi's monthly sales.

Intrigued, Ravi asked, "So you're telling me there's a way to calculate what happens next, just by looking at how things changed before?" Meera nodded, "Exactly! And once you learn how to solve these equations, it feels like having a small window into the future."

This unit is your version of Meera's explanation to Ravi. It is your chance to understand how we track and predict changes over time using simple rules and patterns. Just like Ravi, you might find that these ideas help make sense of many real-life situations - not just in economics, but in life and planning too.

Keywords

First Order Difference Equation, Linear Difference Equation, Homogeneous Equation, Non-Homogeneous Equation, General Solution, Particular Solution

Discussion

2.2.1 Difference Equation

A difference equation expresses a relationship between a dependent variable and a lagged independent variable (or variables) which changes at discrete intervals of time

The order of a difference equation is determined by the greatest number of periods lagged. A first-order difference equation expresses a time lag of one period; a second-order, two periods; etc.

The change in y as t changes from t to $t+1$ is called the first difference of y.

for example, $Q_x = a + bP_{t-1}$, is the difference equation of first order.

$I_t = c(Y_{t-1} - Y_{t-2)}$ is the difference equation of second order

It is written $\frac{\Delta y}{\Delta t} = \Delta y_t - y_{t+1} - y_t$ where Δ is an operator replacing $\frac{d}{dt}$ that is used to measure continuous change in differential equations. The solution of a difference equation defines y for every value of t and does not contain a difference expression.

Solution of first-order difference equation

Iterative Method

Let the difference equation is $y_{t+1} = by_t$ with initial condition of y is y0.

Put $t = 0,\ \ y_1 = by_0$

Put $t = 1,\ \ y_2 = by_1 = b.by_0 = b^2y_0$

Put $t = 2,\ \ y_3 = by_2 = b.b^2y_0 = b^3y_0$ etc

$y_n = b^n y_0$ is the solution of the difference equation.

is a recurrence relation that expresses the value of a sequence at the next step $(x_{n+1}$ in terms of its current value x_n.

General Form of a First-Order Difference Equation

Given a linear first-order difference equation

$\boldsymbol{y_t = b\ y_{t-1} + a}$

where a and b are constants, the general formula for a definite solution is

$\boldsymbol{y_t = \left(y_0 - \frac{a}{1-b}\right)\ b^t + \frac{a}{1-b}}$ when $b \neq 1$

$\boldsymbol{y_t = \ y_0 + at}$ when $b = 1$

Illustration.2.2.1

Solve the difference equation $y_t\ = -7y_{t-1} + 16$ and $y_0\ = 5$. Check the answer in the original equation.

Solution

Comparing with the first order equation $\boldsymbol{y_t = b\ y_{t-1} + a,}$

$b = -7,\ \ a = 16.$ Given $y_0\ = 5$

Solution is $\boldsymbol{y_t = \left(y_0 - \frac{a}{1-b}\right)\ b^t + \frac{a}{1-b}}$

$$y_t = \left(5 - \frac{16}{1+7}\right)\ (-7)^t + \frac{16}{1+7}$$

$$= \left(5 - \frac{16}{8}\right)\ (-7)^t + \frac{16}{8}$$

$$= (5-2)\ (-7)^t + 2$$

$$= 3\ (-7)^t + 2$$

To check the answer, put $t = 0, \quad y_0 = 3\ (-7)^0 + 2 = 3 + 2 = 5$

$t = 1, \quad y_1 = 3\ (-7)^1 + 2 = -21 + 2 = -19$

Substituting $y_1\ = -19$ for y_t and $y_0\ = 5$ the original equation

$$-19 = -7 \times 5 + 16 = -35 + 16 = -19$$

Illustration.2.2.2

Solve the difference equation $y_t\ = -\frac{1}{4}y_{t-1} + 60$ and $y_0\ = 8$. Check the answer in the original equation.

Solution

Comparing with the first order equation $\boldsymbol{y_t = b\ y_{t-1} + a,}$

$b = -\frac{1}{4},\ \ a = 60.$ Given $y_0\ = 8$

Solution is $\boldsymbol{y_t = \left(y_0 - \frac{a}{1-b}\right) b^t + \frac{a}{1-b}}$

$$y_t = \left(8 - \frac{60}{1+\frac{1}{4}}\right) (-\frac{1}{4})^t + \frac{60}{1+\frac{1}{4}}$$

$$= \left(8 - \frac{60 \times 4}{5}\right) (-\frac{1}{4})^t + \frac{60 \times 4}{5}$$

$$= (8 - 48)\ (-\frac{1}{4})^t + 48$$

$$= -40\ (-\frac{1}{4})^t + 48$$

To check the answer, put $t = 0, \quad y_0 = -40\left(-\frac{1}{4}\right)^0 + 48 = -40 + 48 = 8$

$t = 1, \quad y_1 = -40\ (-\frac{1}{4})^1 + 48 = 10 + 48 = 58$

Substituting in the original equation,

$y_1 = -\frac{1}{4} y_0 + 60$, $58 = -\frac{1}{4} \times 8 + 60$

$$= -2 + 60 = 58$$

The solution to the difference equation $y_t = -\frac{1}{4} y_{t-1} + 60$ with the initial condition $y_0 = 8$ is:

$$y_t = -40\left(-\frac{1}{4}\right)^t + 48$$

This solution satisfies both the initial condition and the original difference equation.

Illustration 2.2.3

Solve the difference equation $y_t = 6y_{t-1}$. Check the answer using $t = 0$ and $t = 1$.

Solution

Comparing with the first order equation $\boldsymbol{y_t = b\ y_{t-1} + a}$,

$b = 6, \quad a = 0.$

Solution is $y_t = (y_0 - 0)\ 6^t$

$y_t = y_0 6^t = A6^t$ where $A = y_0$

At $t = 0, \quad y_0 = A6^0 = A$

At $t = 1, \quad y_1 = A6^1 = 6A$

Substituting in the original equation,

$y_1 = 6y_0 = 6A$,

The solution to the difference equation $y_t = 6y_{t-1}$ is: $y_t = A6^t$

where $A = y_0$. This solution satisfies both the initial condition and the original difference equation.

Illustration.2.2.4

Solve the difference equation $y_t = y_{t-1} - 25$ and $y_0 = 40$.

Solution

Comparing with the first order equation $\boldsymbol{y_t = b\, y_{t-1} + a}$,

$b = 1 \quad a = -25$. Given $y_0 = 40$

Solution is $\boldsymbol{y_t = y_0 + at}$ when $b = 1$

$y_t = (40 - 25t)$

To check the answer, put $t = 0, \quad y_0 = 40 - 25 \times 0 = 40$

$t = 1, \quad y_1 = 40 - 25 = 15$

Substituting in the original equation,

$y_1 = y_0 - 25$, $15 = -40 - 25 = 15$

Illustration 2.2.5

Solve the difference equation $5y_t + 2y_{t-1} - 140 = 0$ and $y_0 = 40$. Check the answer in the original equation.

Solution

$5y_t + 2y_{t-1} - 140 = 0$

$$y_t = -\frac{2}{5}y_{t-1} + 28$$

Comparing with the first order equation $\boldsymbol{y_t = b\, y_{t-1} + a}$,

$b = -\frac{2}{5}$, $a = 28$. Given $y_0 = 40$

Solution is $\boldsymbol{y_t = \left(y_0 - \frac{a}{1-b}\right) b^t + \frac{a}{1-b}}$

$$y_t = \left(40 - \frac{28}{1 + \frac{2}{5}}\right)\left(-\frac{2}{5}\right)^t + \frac{28}{1 + \frac{2}{5}}$$

$$= \left(40 - \frac{28 \times 5}{7}\right)\left(-\frac{2}{5}\right)^t + \frac{28 \times 5}{7}$$

$$= (40 - 20)\left(-\frac{2}{5}\right)^t + 20$$

$$= 20 \times \left(-\frac{2}{5}\right)^t + 20$$

To check the answer, put $t = 0$, $y_0 = 20 + 20 = 40$

$t = 1$, $y_1 = 20 \times -\frac{2}{5} + 20 = -8 + 20 = 12$

Substituting in the original equation,

$$y_1 = -\frac{2}{5}y_0 + \frac{140}{5} = -\frac{2}{5} \times 40 + \frac{140}{5} = -16 + 28 = 12$$

Summarised Overview

A difference equation expresses a relationship between a dependent variable and a lagged independent variable (or variables) which changes at discrete intervals of time

General Form of a First-Order Difference Equation $y_t = b\, y_{t-1} + a$.

General solution is $y_t = \left(y_0 - \frac{a}{1-b}\right) b^t + \frac{a}{1-b}$ when $b \neq 1$

Assignments

1. Solve the difference equation $y_t = \frac{1}{8}y_{t-1}$. Check the answer in the original equation.
2. Solve the difference equation $y_t + 3y_{t-1} + 8 = 0$ and $y_0 = 16$. Check the answer in the original equation.
3. Solve the difference equation $y_t - y_{t-1} = 17$
4. Solve the difference equation $\Delta y_t = y_t + 13$ and $y_0 = 45$. Check the answer in the original equation.

References

1. Edward T Dowling, Schaum's Outline Series (2001), *Introduction to Mathematical Economics*, Third Edition, McGRAW-HILL

2. Sreenath Baruah(2012): *Basic Mathematics and its applications in Economics*, Macmillan India Ltd.

3. Mehta and Madnani (2022): *Mathematics for Economists*, Sultan Chand & Sons, New Delhi

4. Monga G S (2014), *Mathematics and Statistics for Economics*, Vikas Publishing House Pvt Ltd

Suggested Readings

1. Yamane, Taro. (2012). *Mathematics for Economists: An Elementary Survey*. New Delhi: Prentice Hall of India.

2. Chiang, A.C. (2008), *Fundamental Methods of Mathematical Economics*, McGraw Hill, New York.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

3

UNIT

Applications of Difference Equations in Economics – Dynamic Stability Conditions

Learning Outcomes

After completing this unit, the learner will able to

- understand the structure and role of difference equations in economic modeling
- identify and interpret the order and degree of difference equations using appropriate methods
- analyse the dynamic stability of difference equations using the characteristic root

Background

In economics, mathematical tools and techniques are essential for analysing how economic variables evolve over time. When these changes occur at specific, discrete intervals - such as quarterly income reports, annual investments, or periodic policy implementations - difference equations provide a powerful framework for modeling and analysis. These equations enable economists to track the evolution of variables from one time period to the next, making them especially relevant for understanding real-world economic behavior in discrete time settings. Difference equations are widely applied in dynamic economic modeling, where they capture intertemporal relationships such as income and consumption over time, investment planning, and the effects of economic policy across successive periods. Unlike differential equations that model continuous change, difference equations are tailored for discrete steps, offering a more realistic representation of many economic processes. One of the critical concepts in studying difference equations is dynamic stability - the idea of whether an economic system returns to its equilibrium path following a disturbance. Analyzing stability conditions allows economists to evaluate whether policy interventions or market shocks lead to convergence back to equilibrium, divergence away from it, or cyclical fluctuations. This is crucial for developing reliable forecasts and designing stable economic systems.

Keywords

Difference Equation, Dynamic Stability, Lagged Variables, Order and Degree, Oscillation, Complementary Function

Discussion

2.3.1 Applications of Difference Equations in Economics

In dynamic analysis, sometimes we may consider the explanatory variable or the time variable as discrete instead of continuous. When we deal with discrete time, the dependent variable y, undergoes a change only when the time variable t, changes from one integer value to the next such as from t = 1 to t = 2. Such discrete version of dynamic analysis is also known as period-analysis. There is no such specific period for the time variable. It may be a year, or a month or a week or a day depending on the nature of the dynamic model. A period is defined as a length of time that elapses before the variable y undergoes a change.

In many economic studies the focus is not on relating continuous changes of variables but is on discrete changes. In planning models, for example, the comparison is between the initial base year and the terminal year and change in investment over the period is related to change in time over a period. Both the changes are said to be discrete.

Let Y(t) be a function of t, The change in Y(t) due to a change in t by a positive number h is called the first forward difference of Y_t, relative to the increment h, and is denoted by Δy_t.

Thus,

$$\Delta y_t = Y\ (t+h) - Y(t)$$

In economic applications t generally denotes time and h is equal to one period. Y(t) is also written as Y_t.

A difference equation relates the independent variable, the dependent variable and its successive differences.

A difference equation expresses a relationship between a dependent variable and a lagged independent variable (or variables) which changes at discrete intervals of time, for example, $I_t = f\ (Y_{t-1})$, where I and Y are measured at the end of each year. The order of a difference equation is determined by the greatest number of periods lagged. A first-order difference equation expresses a time lag of one period; a second-order, two periods; etc. The change in y as t changes from t to t + 1is called the first difference of y. It is written

$$\frac{\Delta y}{\Delta t} = \Delta y_t = y_{t+1} - y_t$$

where, Δ is an operator replacing $\frac{d}{dt}$ that is used to measure continuous change in differential equations. The solution of a difference equation defines y for every value of t and does not contain a difference expression.

In economics, we often come across situations when variables occur with discrete time lags. We now consider the calculus of finite differences which deals with this subject. The use of differences as variables amounts to using lagged variables corresponding to previous time periods.

The difference $\Delta y_t = y_{t+1} - y_t$ is called the first difference of y_t and provides the rule for computing Δy_t. It may be noted that y is a function of t but that it takes only integral values. Thus y can be evaluated at intervals and not continuously.

The second difference is written

$$\Delta^2 y_t = y_{t+2} - 2y_{t+1} + y_t$$

$$\Delta^3 y_t = y_{t+3} - 3y_{t+2} + 3y_{t+1} - y_t$$

and so on

An ordinary difference equation is an equation involving differences like Δy_t, $\Delta^2 y_t$, etc.

The order of a difference equation is the highest differences in the equation. The degree of a difference equation is the degree of the term of the highest order. A difference equation of degree one is said to be linear.

Example.1 : Each of the following is a difference equation of the order indicated

$I_t = a\,(y_{t-1} - y_{t-2})$	order 2
$Q_s = a + bP_{t-1}$	order 1
$y_{t+3} - 7y_{t+2} + 4y_{t+1} + 8y_t = 10$	order 3
$\Delta y_t = 3y_t$	order 1

Substituting from $\frac{\Delta y}{\Delta t} = \Delta y_t = y_{t+1} - y_t$

$y_{t+1} - y_t = 3y_t$

$y_{t+1} = 4y_t$ order 1

Example.2 : Given that the initial value of y is $y_{0,}$ in the difference equation

$$y_{t+1} = by_t$$

A solution is found as follows. By successive substitutions of t = 0, 1, 2, 3 etc in $y_{t+1} = by_t$

If t = 0, $\rightarrow y_{0+1} = by_0 \rightarrow y_1 = by_0$

If t = 1, → $y_{1+1} = by_1$ → $y_2 = by_1$ → $b(by_0) = b^2y_0$

If t = 2, → $y_{2+1} = by_2$ → $y_3 = by_2$ → $b(b^2y_0) = b^3y_0$

If t = 3, → $y_{3+1} = by_3$ → $y_4 = by_3$ → $b(b^3y_0) = b^4y_0$

Thus, for any period t,

$$y_t = b^t y_0$$

This method is called the iterative method. Since y_0 is a constant, notice the crucial role b plays in determining values for y as t changes.

General Formula for First-Order Linear Difference Equations

We observed that solution of a difference equation is similar to that of a differential equation. The general solution of the difference equation also consists of two components, a particular integral, Y_p and a complementary function $Y_{c.}$ The Y_p component will represent the equilibrium level of Y and Y_c component will depict deviations of the time path from the equilibrium level. The sum of Y_p and Y_c will give the general solution of any difference equation. A function is a solution of a difference equation if it makes the equation a true statement and it satisfies the equation. If the solution satisfies the equation without reference to initial conditions, it is called general solution. If, in addition, it satisfies the initial conditions, it is termed as a particular solution.

Illustration 2.3.1

Show that $Y_t = t + C$ is a solution of $Y_{t+1} - Y_t = 1$. Also find a particular solution if $Y_0 = 3$ when t = 0

Solution

Since $Y_t = t + C$ is the solution of the difference equation :

$$Y_{t+1} - Y_t = 1$$

The equation should get satisfied on substituting the values of Y_{t+1} and $Y_{t\,.}$

Given : $Y_t = t + C$

$$Y_{t+1} = (t+1) + C$$

Substituting the above two equations in $Y_{t+1} + Y_t = 1$ we get

$$(t+1) + C - (t + C) = 1$$

Hence, the equation is satisfied, therefore $Y_t = t + C$ is the general solution of the given equation.

Initial condition: $Y_0 = 3$

Substituting in the solution: 3 = 0 + C , or C = 3

So, the particular solution becomes: $Y_t = t + 3$

Illustration 2.3.2

Solve the equation Δy_t = -0.3y_t

Solution

Y_{t+1} - Y_t = -0.3y_t

Y_{t+1} - 0.7Y_t = 0

Y_{t+1} = 0.7Y_t

When t = 0 Y_1 = 0.7Y_t

t = 1, $Y_2 = 0.7Y_1 = 0.7\,(0.7Y_0) = (0.7)^2\,Y_0$

t = 2, $Y_3 = 0.7Y_2 = 0.7\,(0.7)^2Y_0) = (0.7)^3\,Y_0$

t = 3, $Y_4 = 0.7Y_3 = 0.7\,(0.7)^3Y_0) = (0.7)^4\,Y_0$

Generalising : $Y_t = 0.7Y_{t-1} = 0.7\,(0.7)^{t-1}Y_0) = (0.7)^t\,Y_0$

Thus, the solution of the given equation will be

$Y_t = Y_0\,(0.7)^t$

Given a first-order difference equation which is linear (i.e., all the variables are raised to the first power and there are no cross products)

$y_t = by_{t-1} + a$

where b and a are constants, the general formula for a definite solution is

$y_t = (y_0 - \frac{a}{1-b})\,b^t + \frac{a}{1-b}$ when b ≠ 1

$y_t = y_0 + at$ when b = 1

If no initial condition is given, an arbitrary constant A is used for $y_0 - \frac{a}{1-b}$ in $y_t = (y_0 - \frac{a}{1-b})\,b^t + \frac{a}{1-b}$ equation and for y_0 in $y_t = y_0 + at$. This is called a general solution.

Illustration 2.3.3

Consider the difference equation $y_t = -7y_{t-1} + 16$ and $y_0 = 5$. In the equation, b = -7 and a = 16. since b ≠ 1

Solution

It is solved by using $y_t = (y_0 - \frac{a}{1-b})\,b^t + \frac{a}{1-b}$

$$y_t = (5 - \frac{16}{1+7})\,(-7)^t + \frac{16}{1+7} = 3(-7)^t + 2$$

To check the answer, substitute t = 0 and t = 1 in the above

$y_0 = 3(-7)^0 + 2 = 5$ since $(-7)^0 = 1$

$y_1 = 3(-7)^1 + 2 = -19$

Substituting $y_1 = -19$ for y_t and $y_0 = 5$ for y_{t-1} in the original equation,

-19 = -7(5) + 16 = -35 + 16 = -19

2.3.2 Dynamic Stability Conditions

Equation $y_t = \left(y_0 - \frac{a}{1-b}\right)b^t + \frac{a}{1-b}$ can be expressed in the general form

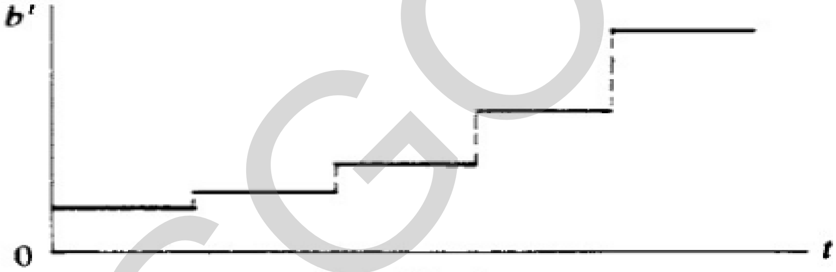
$$y_t = Ab^t + c$$

Where,

$$A = y_0 - \frac{a}{1-b}$$

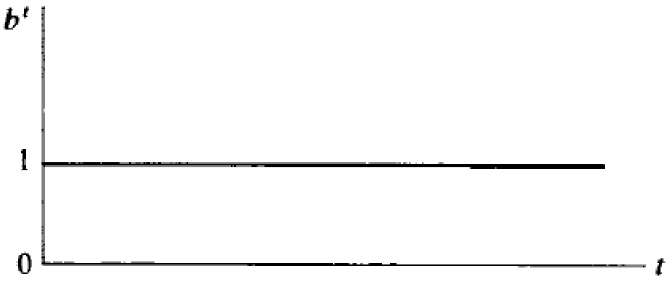
$$c = \frac{a}{1-b}$$

Here Ab^t is called the complementary function and c is the particular solution. The particular solution expresses the intertemporal equilibrium level of y; the complementary function represents the deviations from that equilibrium. The equation $y_t = Ab^t + c$ will be dynamically stable, therefore, only if the complementary function $Ab^t \rightarrow 0$, as $t \rightarrow \infty$. All depends on the base b. Assuming A = 1 and c = 0 for the moment, the exponential expression b^t will generate seven different time paths depending on the value of b, as illustrated in the following figure. In the equation $y_t = b^t$, b can range from $-\infty$ to ∞.



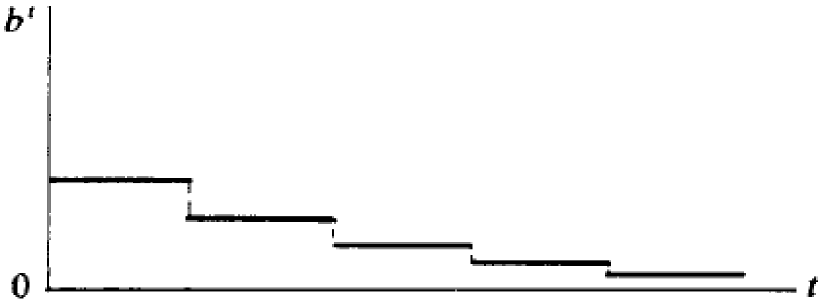
(a) $b > 1$

If $|b| > 1 \rightarrow$ The time path will explode and move farther and farther away from equilibrium. If $b > 1$, b^t increases at an increasing rate as t increases, thus moving farther and farther away from the horizontal axis, which is a step function representing changes at discrete intervals of time, not a continuous function. Assume b = 3. Then as t goes from 0 to 4, b^t = 1, 3, 9, 27, 81.



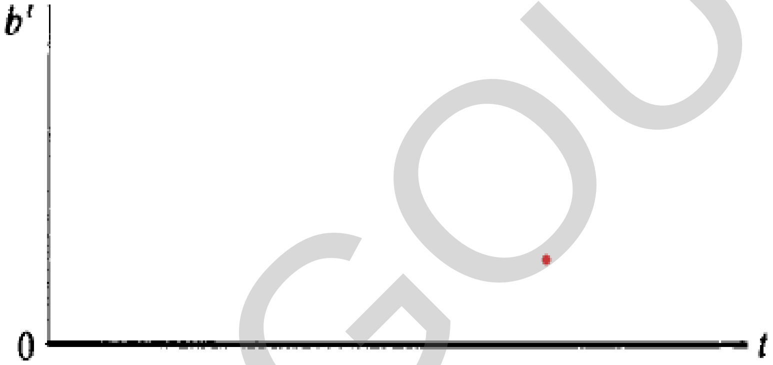
(b) $b = 1$

If b =1, $b^t = 1$ for all values of t. This is represented by a horizontal line.



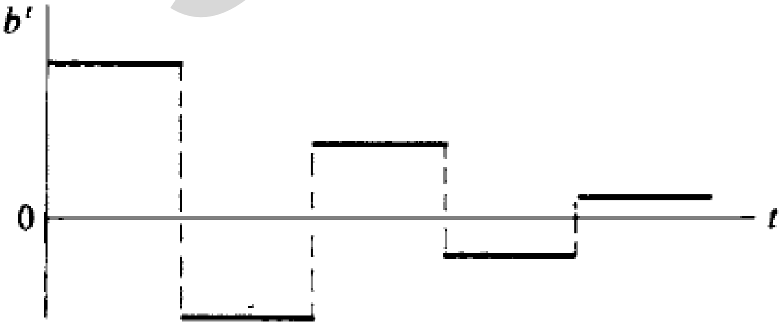
(c) $0 < b < 1$

If $|b| < 1$ → The time path will be damped and move toward equilibrium. If $0 < b < 1$, then b is positive fraction and b^t decreases as t increases, drawing closer and closer to the horizontal axis, but always remaining positive. Assume b = 1/3. Then as t goes from 0 to 4, b^t = 1, 1/3, 1/9, 1/27, 1/81.



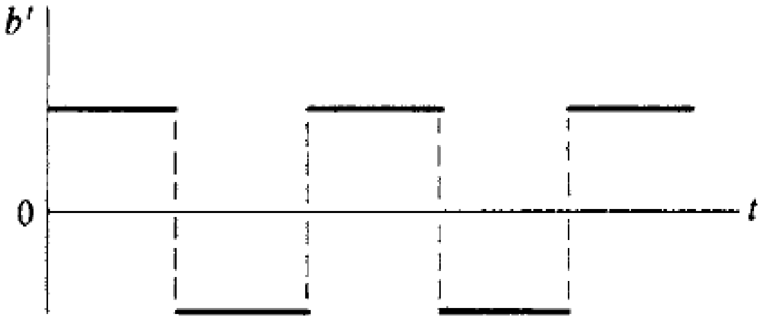
(d) $b = 0$

If b = 0, then $b^t = 0$ for all values of t



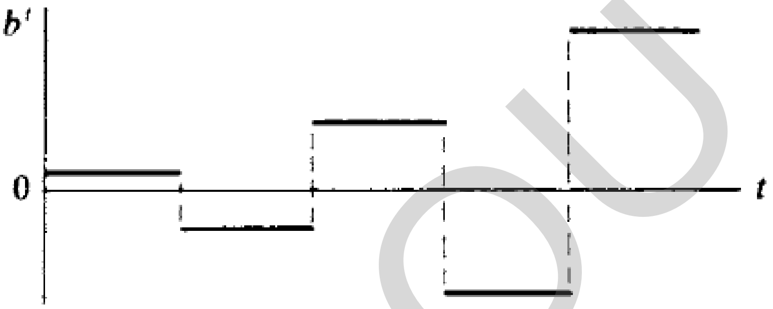
(e) $-1 < b < 0$

If $b < 0$, the time path will oscillate between positive and negative values; if $b > 0$, the time path will be non oscillating. If $-1 < b < 0$, then b is a negative fraction; b^t will alternate in sign and draw closer and closer to the horizontal axis as t increases. Assume b = -1/3. Then as t goes from 0 to 4, b^t = 1, -1/3, 1/9, -1/27, 1/81.



(*f*) $b = -1$

If b = -1, then b^t oscillates between +1 and -1



(*g*) $b < -1$

If b < -1, then b^t will oscillate and move farther and farther away from the horizontal axis. Assume b = -3. Then b^t = 1, -3, 9, -27, 81 and t goes from 0 to 4.

If A ≠ 1, the value of the multiplicative constant will scale up or down the magnitude of b^t, but will not change the basic pattern of movement. If A = -1, a mirror image of the time path of b^t with respect to the horizontal axis will be produced. If c ≠ 0, the vertical intercept of the graph is affected, and the graph shifts up or down accordingly.

In short, if

$\|b\| > 1$	→	The time path explodes
$\|b\| < 1$	→	The time path converges
b > 0	→	The time path is non oscillating
b < 0	→	The time path oscillates

Example.: In the equation $y_t = 6\left(-\frac{1}{4}\right)^t + 6$, since $b = -\frac{1}{4} < 0$, the time path is oscillates. Since $|b| < 1$, the time path converges

When $y_t = 5\ (6)^t + 9$, and b = 6 > 0, there is no oscillation. With $|b| > 1$, the time path explodes.

Illustration 2.3.4

Solve the difference equation $y_t = 8y_{t-1}$

Solution

a. Here, b = 8 and a = 0.

Using $y_t = (y_0 - \frac{a}{1-b}) b^t + \frac{a}{1-b}$ when $b \neq 1$

$y_t = (y_0 - 0) (8)^t + 0 = y_0(8)^t =$ $A(8)^t$

Where A, as more generally used unspecified constant, replaces y_0

b. Estimating $A(8)^t$ at t = 0 and t = 1

$y_0 = A(8)^0 = A$

$y_1 = A(8)^1 = 8A$

Substituting $y_0 = A$ for y_{t-1} and $y_1 = 8A$ for y_t in the original problems, 8A = 8(A).

c. With the base b = 8 in the above equation is positive and greater than 1, that is, $b > 0$ and $|b| > 1$, the time path is nonoscillating and explosive.

Illustration 2.3.5

Solve the difference equation $y_t = -\frac{1}{4} y_{t-1} + 60$ and $y_0 = 8$

Solution

a. Here, $b = -\frac{1}{4}$ and a = 60.

Using $y_t = (y_0 - \frac{a}{1-b}) b^t + \frac{a}{1-b}$ when $b \neq 1$

$y_t = (8 - \frac{60}{1+\frac{1}{4}}) (-\frac{1}{4})^t + \frac{60}{1+\frac{1}{4}} = -40 (-\frac{1}{4})^t + 48$

Where A, as more generally used unspecified constant, replaces y_0

b. Estimating $-40 (-\frac{1}{4})^t + 48$ at t = 0 and t = 1

$y_0 = -40 (-\frac{1}{4})^t + 48 = 8$

$y_1 = -40 (-\frac{1}{4})^t + 48 = 58$

Substituting $y_0 = 8$ for y_{t-1} and $y_1 = 58$ for y_t in the original problems, $y_t = -\frac{1}{4} y_{t-1} + 60$.

$58 = -\frac{1}{4} (8) + 60 = 58$

c. With the base $b = -\frac{1}{4}$ in the above equation is negative and less than 1, that is, $b < 0$ and $|b| < 1$, the time path oscillates and converges.

Summarised Overview

This section explores how difference equations help model economic variables that evolve at discrete intervals. Students learn about the forward difference operator (Δ), the formulation of difference equations, and their application in lag-based economic models. The note emphasizes first-order linear difference equations and outlines the general and particular solutions. Further, dynamic stability is analysed by examining the base (b) in expressions like $y_t = Ab^t+c$ Depending on the value of b, systems may converge to equilibrium, diverge, or oscillate. Real-world economic implications, such as investment planning and growth modeling, are linked to these mathematical outcomes.

Assignments

1. Define difference equation and explain its economic relevance.
2. What is the significance of the order and degree in a difference equation?
3. How do you determine whether a time path is convergent, divergent, or oscillatory?
4. Solve the difference equation $y_{t+1}=0.5y_t$ given $y_0=10$.
5. In the equation $y_t = A(-0.4)^t + 5$, what does the time path suggest about stability and behavior.
6. What are the conditions for dynamic stability in a first-order linear difference equation?
7. Given the equation $y_{t+1} = -0.5y_t+10$, derive the general and particular solutions. Discuss the behavior of the solution over time.
8. Plot the time paths for $y_t= (0.5)^t$, $y_t=(-0.5)^t$ and $y_t=(1.5)^t$ using assumed values of y_0. Comment on their economic interpretations.
9. Discuss the economic implications of dynamic instability in a macroeconomic model.
10. Solve the first-order difference equation and analyze its stability: $x_{t+1} = 0.7x_t + 10$, with $x_0 = 50$.
11. Consider the equation $K_{t+1} = 1.1k_{t-5}$.

12. Find the general solution.

13. Check whether the equilibrium is stable or unstable.

References

1. Edward T Dowling, Schaum's Outline Series (2001), Introduction to Mathematical Economics, Third Edition, McGRAW-HILL
2. Yamane, Taro. (2012). Mathematics for Economists: An Elementary Survey. New Delhi: Prentice Hall of India.
3. Sreenath Baruah(2012): Basic Mathematics and its applications in Economics, Macmillan India Ltd.
4. Mehta and Madnani (2022): Mathematics for Economists, Sultan Chand & Sons, New Delhi
5. Monga G S (2014), Mathematics and Statistics for Economics, Vikas Publishing House Pvt Ltd

Suggested Readings

1. Chiang, A.C. (2008), Fundamental Methods of Mathematical Economics, McGraw Hill, New York.
2. Y.P. Agarwal: Statistical Methods: Concepts, Application and Computation, Sterling Publishers 1986
3. Hooda R.P: Statistics for Business and Economics , Mac Million, New Delhi

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

UNIT 4

Cobb-Web and Harrod model-Lagged Income Determination Model

Learning Outcomes

After completing this unit, the learner will able to:

- understand the formulation and types of difference equations in economic modelling
- identify and interpret the economic applications of stable and unstable dynamic systems of difference equations using appropriate methods
- analyse the stability concepts to real-world economic scenarios

Background

In economics, dynamic models often explore how agents respond to delayed information or adjustment lags, leading to fluctuations in variables like prices, output, and income. Difference equations play a crucial role in capturing these time-lagged relationships, allowing economists to analyze the path of economic variables as they adjust over successive periods. The classical examples of such dynamic models are the Cobb-Web model, Harrod model and the Lagged Income Determination model.

The Cob-Web model provides insight into markets where production decisions are based on past prices, which leads to a time lag between supply decisions and actual market outcomes. This model is particularly applicable to agricultural markets, where producers decide how much to supply based on previous season's prices. The interaction between lagged supply and current demand creates a system of difference equations that describe the price-output dynamics over time. The model's analysis reveals conditions under which prices converge to equilibrium, diverge, or oscillate indefinitely making it a foundational case of dynamic stability analysis.

The Harrod model incorporates time lags in income determination by emphasizing the relationship between planned savings and investment over discrete time periods. Central to this model is the assumption that output (or income) adjusts with a lag in response to planned investment decisions, resulting in cyclical or explosive growth trajectories depending on the parameters involved. By modeling national income across successive periods, the Harrod model highlights how instability can arise in capitalist economies due to the mismatch between expected and actual outcomes.

The Lagged Income Determination Model utilizes difference equations to represent how current income levels depend on investment decisions and savings behavior from previous periods. This model captures the realistic time lag between planned investment and the actual adjustment of output or income in an economy. The lagged response reflects the inertia in economic systems where changes in aggregate demand, production, or capital formation do not occur instantaneously. Through this structure, the model reveals important insights about the stability or instability of economic growth paths, particularly how small deviations from equilibrium can lead to cyclical fluctuations or even explosive growth or decline. By applying mathematical techniques to examine the dynamic behavior of income over discrete time intervals, the Lagged Income Determination Model offers a foundational approach to studying growth dynamics and policy impacts in a time-dependent economic environment.

Keywords

Difference Equation, Lagged income, Dynamic Stability, Income Determination, Time Lag, Harrod Model, Cob-Web Model

Discussion

2.4.1 Cob-Web and Harrod model

Cobweb means spider's web generally when old and dusty. The Cobweb Theory was first coined by Nicholas Kaldor in 1934. However, according to literature, Henry Schultz, along with other economists, such as Jan Tinbergen and Althus Hanau are generally associated with the Cobweb Model.

The Cobweb Theorem is based on time lag between the supply and demand adjustments due to price fluctuations. When a farmer goes to the market with small supply and if the demand is greater than the supply, this raises the price in the market. If this high price continues in the next period, the farmer will produce more and will go to the market with high supply in the next period. Now, consumers prefer to buy at low cost and if the prevailing or current price in the market is high, demand will be less. This will give rise to a situation of surplus and the price falls in the market. If the farmer expects a low price in the next period, the production will fall and the supply will be less, resulting in high price again. This cycle continues for successive periods. Here,the quantity demanded depends on the current price and quantity supplied depends on the previous period price.

In agriculture, time lag is observed between planting and harvesting, and in this time lag, exists a rise and fall in prices that results in the adjustments of demand and supply and this cycle gives rise to a web-like structure. Hence this theory was termed cobweb theory

The Cobweb model, a dynamic model of supply and demand, can be effectively analysed using first-order difference equations. The model demonstrates how price fluctuations and quantity adjustments can create a web-like pattern over time, particularly in markets with lags between planting and harvesting. In a market situation where the output decision of producer depends on the price of the previous period and the demand for the product being function of current period price, such a market model is popularly known as Cobweb model. Such a market model is very appropriate for agricultural product or for the product of perishable commodities.

For many products, such as agricultural commodities, which are planted a year before marketing, current supply depends on last year's prices. This poses interesting stability questions.

If

$$Q_d(t) = c + bP_t \qquad \text{and} \qquad Q_s(t) = g + hP_{t-1}$$

In equilibrium,

$$c + bP_t = g + hP_{t-1}$$

$$bP_t = hP_{t-1} + g - c$$

Dividing $bP_t = hP_{t-1} + g - c$ by b to confirm to $y_t = by_{t-1} + a$

$$P_t = \frac{h}{b} P_{t-1} + \frac{g-c}{b}$$

Since b < 0 and h > 0 under the normal demand and supply conditions,

h/b ≠ 1 using $y_t = (y_0 - \frac{a}{1-b}) b^t + \frac{a}{1-b}$

$$P_t = \left[P_0 - \frac{(g-c)/b}{1-h/b}\right] (\frac{h}{b})^t + \frac{(g-c)/b}{1-h/b}$$

$$= \left(P_0 - \frac{g-c}{b-h}\right) (\frac{h}{b})^t + \frac{g-c}{b-h}$$

When the model is in equilibrium, $P_t = P_{t-1}$. Substituting P_e for P_t and P_{t-1} in $c + bP_t = g + hP_{t-1}$

$$P_e = \frac{g-c}{b-h}$$

Substituting in $\left(P_0 - \frac{g-c}{b-h}\right)\left(\frac{h}{b}\right)^t + \frac{g-c}{b-h}$

$$P_t = (P_0 - P_e)\left(\frac{h}{b}\right)^t + P_e$$

The nature of the time path depends on $(\frac{h}{b})$, which is the ratio of the slopes of the supply and demand curves respectively. There may be three cases.

With an ordinary negative demand function and positive supply function, b < 0 and h > 0. Therefore, h/b < 0 and the time path will oscillate.

Case 1 : → If $|h| > |b|$, $|h/b| > 1$, and the time path P_t explodes

Case 2 : → If $|h| = |b|$, h/b = -1, and the time oscillates uniformly

Case 3 : → If $|h| < |b|$, $|h/b| < 1$, and the time path converges and P_t approaches P_e

Case 1 : → If $|h| > |b|$, $|h/b| > 1$, and the time path P_t explodes

If the ratio ($|h/b| > 1$) , this means that the slope of the demand curve, that is b, is less than the slope of the supply curve, that is h . In other words, the slope of the supply curve is steeper than that of the demand curve.

If ($|h/b| > 1$) , and the value of t goes on increasing, the ratio $(\frac{h}{b})^t$ goes on increasing. If t → ∞, $(\frac{h}{b})^t \rightarrow \infty$ and the expression $(P_0 - P_e)(\frac{h}{b})^t$ will tend to infinity. Therefore, $P_t \neq P_e$.

This shows that if the slope of the supply curve is greater than the slope of the demand curve, the time path of price moves away from the equilibrium price and is said to be divergent or market is dynamically unstable.

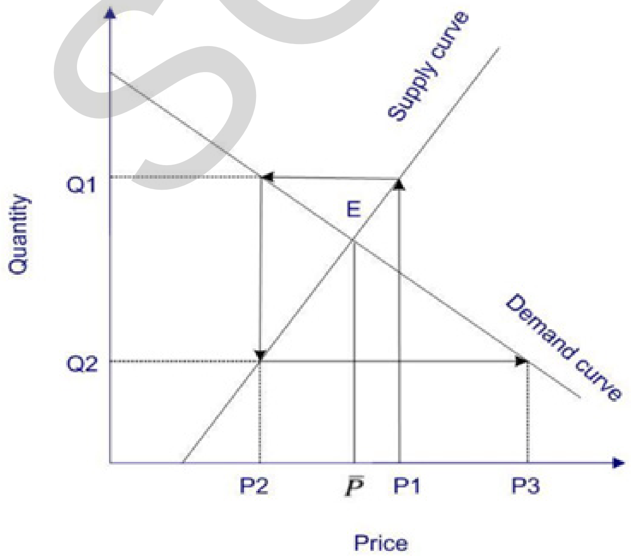


Fig 2.4.1 Flatter Demand and Steeper Supply Curve

As can be seen in the above figure, the demand is flatter than supply, there is an unstable equilibrium. If the slope of the supply curve is steeper than that of the demand curve, assuming that the current price is greater than the market price, the time path of price is divergent or moving away from the equilibrium price. If P1 is the initial price, P3 is greater than P1, which indicates that P1 diverges from $\bar{P}$ (P_e).

Case 2 : → If $|h|$ = $|b|$, h/b = -1, and the time oscillates uniformly

If the ratio h/b = 1, this means that the slope of the demand curve and the slope of the supply curve is equal .

If h/b = 1, the ratio $(\frac{h}{b})^t$will be positive or negative, depending on whether t is odd or even. If t is odd, we have -1 and if t is even, we have +1. So, the expression $(P_0 - P_e)\ (\frac{h}{b})^t$ will alternate between -1 and +1 and the time path will be a regular one.

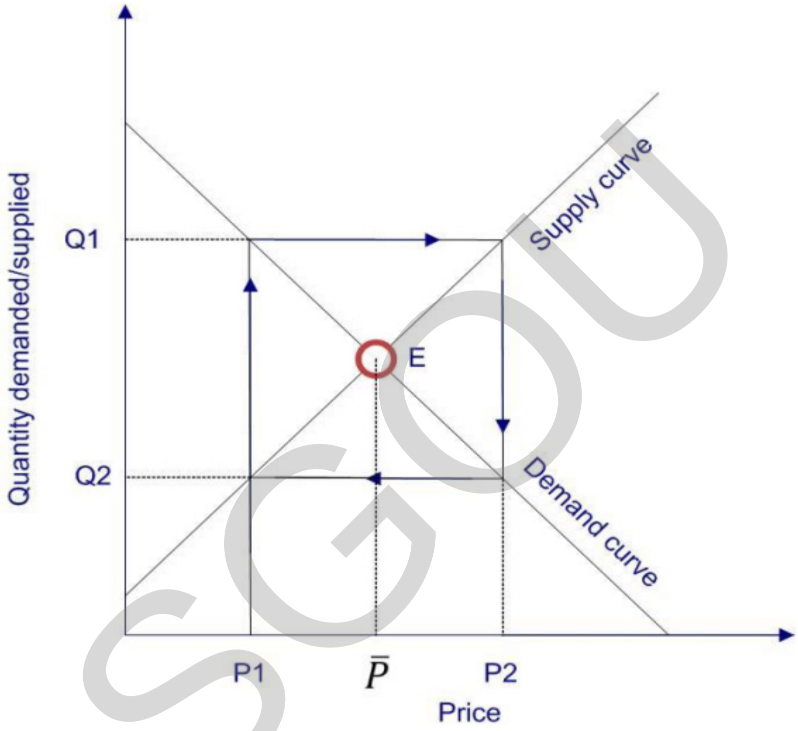


Fig 2.4.2 Equal slope for Demand and Supply Curve

As can be seen in the above figure, the demand and supply having equal slopes and there is an uniform oscillations. The divergence of the current price from the equilibrium price will remain same for different time periods, thus giving rise to a regular time path. The arrows move around the same square.

Case 3 : → If $|h| < |b|$, $|h/b| < 1$, and the time path converges and P_t approaches P_e

If the ratio ($|h/b| < 1$) , this means that the slope of the demand curve, that is b, is greater than the slope of the supply curve, that is h . In other words, the slope of the demand curve is steeper than that of the supply curve.

If ($|h/b| < 1$), and the value of t goes on increasing, the ratio $(\frac{h}{b})^t$ goes on decreasing. If t → ∞, $(\frac{h}{b})^t \rightarrow 0$ and the expression $(P_0 - P_e)\ (\frac{h}{b})^t$ will tend to zero. Therefore, $P_t = P_e$.

This shows that if the slope of the demand curve is greater than the slope of the supply curve, the time path of price moves towards the equilibrium price and is said to be convergent or market is dynamically stable.

As can be seen in the below figure, the supply is flatter than demand, there is a stable equilibrium. If the slope of the demand curve is steeper than that of the supply curve, assuming that the current price is greater than the market price, the time path of price is convergent or moving towards the equilibrium price. If P1 is the initial price, P3 is less than P1, which indicates that P1 $\rightarrow \bar{P}$ (P_e).

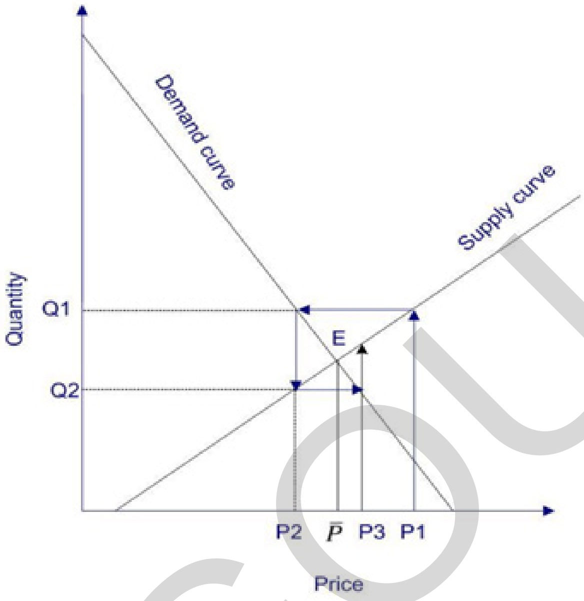


Fig 2.4.3 Steeper Demand and Flatter Supply Curve

In short, when Q = f (P) in supply and demand analysis and the supply curve must be flatter than the demand curve for stability. But if P = f(Q), the demand curve must be flatter or more elastic than the supply curve if the model is to be stable.

Example. 1: Given $Q_{dt} = 100 - 0.5P_t$ and $Q_{st} = 20 + 0.3P_{t-1}$, the market price P_t for any time period and the equilibrium pric P_e can be found as follows. Equating demand and supply,

$$100 - 0.5P_t = 20 + 0.3P_{t-1}$$

$$-0.5P_t = 0.3P_{t-1} - 80$$

Dividing through by -0.5 to confirm $y_t = by_{t-1} + a$,

$$P_t = -0.6P_{t-1} + 160$$

Using $y_t = (y_0 - \frac{a}{1-b})\, b^t + \frac{a}{1-b}$

$$P_t = (P_0 - \frac{160}{1+0.6})\,(-0.6)^t + \frac{160}{1+0.6}$$

$$= (P_0 - 100\,)\,(-0.6)^t + 100$$

Which can be checked by substituting the appropriate values in

$$\left(P_0 - \frac{g-c}{b-h}\right)\left(\frac{h}{b}\right)^t + \frac{g-c}{b-h}$$

From $P_e = \frac{g-c}{b-h}$, $P_e = \frac{20-100}{-0.5-0.3} = \frac{-80}{-0.8} = 100$

With the base b = -0.6 which is negative and less than 1, the time path oscillates and converges. The equilibrium is stable, and P_t will converge to $P_e = 100$ as t → ∞.

Example. 2: For the data given below, determine (a) the market price P_t in any time period, (b) the equilibrium price P_e, and (c) the stability of the time path.

$Q_{dt} = 160 - 0.6P_t$ and $Q_{st} = -20 + 0.4P_{t-1}$, $P_0 = 190$

a. Equating demand and supply

$$160 - 0.6P_t = -20 + 0.4P_{t-1}$$

$$- 0.6P_t = 0.4P_{t-1} - 180$$

Dividing through by -0.6 and using $y_t = (y_0 - \frac{a}{1-b})\, b^t + \frac{a}{1-b}$

$$P_t = -0.67P_{t-1} + 300 = (190 - \frac{300}{1+0.67})\,(-0.67)^t + \frac{300}{1+0.67} = 10(-0.67)^t + 180$$

b. If the market is in equilibrium, $P_t = P_{t-1}$. Substituting P_e for P_t and P_{t-1} in

$$160 - 0.6P_e = -20 + 0.4P_e$$

$P_e = 180$ (which is the second term on the right hand side of $10(-0.67)^t + 180$

c. With b = -0.67, the time path P_t will oscillate and converge.

Example. 3: For the data given below, determine (a) the market price P_t in any time period, (b) the equilibrium price P_e, and (c) the stability of the time path.

$Q_{dt} = 220 - 0.4P_t$ and $Q_{st} = -30 + 0.6P_{t-1}$, $P_0 = 254$

a. Equating demand and supply

$$220 - 0.4P_t = -30 + 0.6P_{t-1}$$

$$- 0.4P_t = 0.6P_{t-1} - 250$$

Dividing through by -0.4 and using $y_t = (y_0 - \frac{a}{1-b})\, b^t + \frac{a}{1-b}$

$$P_t = -1.5P_{t-1} + 625 = (254 - \frac{625}{1+1.5})\,(-1.5)^t + \frac{625}{1+1.5} = 4(-1.5)^t + 250$$

b. If the market is in equilibrium, $P_t = P_{t-1}$. Substituting P_e for P_t and P_{t-1} in

$$220 - 0.4P_e = -30 + 0.6P_e$$

$P_e = 250$ (which is the second term on the right hand side of $4(-1.5)^t + 250$

c. With b = -1.5, the time path P_t will oscillate and explodes.

2.4.2 Harrod Model

Harrod's Model of economic growth has been developed by Sir R.F. Harrod. Harrod defined three different types of rates of growth:

i. The Warranted rate of growth (G_w) → The rate that will keep investment and savings equal through time, therefore maintaining equilibrium

ii. The Actual rate of growth (G_a) → The increment of total production in any unit period expressed as a fraction of total production

iii. The Natural rate of growth (G_n) → The rate of advance which, the increase in population and technological improvements allow.

There is equilibrium situation with full employment in a period when $G_w = G_a = G_n$

The Harrod model attempts to explain the dynamics of growth in the economy. It assumes

$S_t = sY_t$

where 's' is a constant equal to both the MPS and APS. It also assumes the acceleration principle, i.e., investment is proportional to the rate of change of national income over time

$I_t = a\ (Y_t - Y_{t-1})$

where 'a' is a constant equal to both the marginal and average capital-output ratios. In equilibrium,

$I_t = S_t$. Therefore,

$a\ (Y_t - Y_{t-1}) = sY_t \qquad (a - s)Y_t = aY_{t-1}$

Dividing through by a - s to confirm to $y_t = by_{t-1} + a$, $Y_t = a/(a - s)Y_{t-1}$. Using

$Y_t = (Y_0 - \frac{a}{1-b})\ b^t + \frac{a}{1-b}$ since $a/(a - s) \neq 1$

$Y_t = (Y_0 - 0)\ (\frac{a}{a-s})^t + 0 = (\frac{a}{a-s})^t Y_0$

The stability of the time path thus depends on $\frac{a}{a-s}$.

If $a > s$, the time path of Y_t will be explosive as t increases.

If $a < s$, the time path of Y_t will be oscillatory.

Since a = capital - output ratio, which is normally larger than 1, and since s = MPS which is larger than 0 and less than 1, the base $\frac{a}{a-s}$ will be larger than 0 and usually larger than 1. Therefore, Y_t is explosive but nonoscillating. Income will expand indefinitely, which means it has no bounds.

Example.4 : The warranted rate of growth (i.e., the path the economy must follow to have equilibrium between saving and investment each year) can be found as follows in the Harrod model.

From $Y_t = (Y_0 - 0)\,(\frac{a}{a-s})^t + 0 = (\frac{a}{a-s})^t Y_0$, Y_t increases indefinitely. Income in one period is $\frac{a}{a-s}$ times the income of the previous period.

$$Y_1 = (\frac{a}{a-s})Y_0$$

The rate of growth G between the periods is defined as

$$G = \frac{Y_1 - Y_0}{Y_0}$$

Substituting from $Y_1 = (\frac{a}{a-s})Y_0$

$$G = \frac{[a/(a-s)]Y_0 - Y_0}{Y_0} = \frac{[a/(a-s)-1]Y_0}{Y_0}$$

$$= \frac{a}{a-s} - 1 = \frac{a}{a-s} - \frac{a-s}{a-s} = \frac{s}{a-s}$$

The warranted rate of growth, therefore, is

$$G_w = \frac{s}{a-s}$$

Example.5 : Assume that the Marginal Propensity to Save in the Harrod model is 0.14 and the capital-output ratio is 2.14. To find Y_t

$$Y_t = (Y_0 - 0)\,(\frac{a}{a-s})^t + 0 = (\frac{a}{a-s})^t Y_0$$

$$Y_t = (\frac{2.14}{2.14 - 0.14})^t Y_0 = (1.07)^t Y_0$$

The warranted rate of growth is,

$$G_w = \frac{s}{a-s}$$

$$G_w = \frac{0.14}{2.14 - 0.14} = \frac{0.14}{2} = 0.07$$

Example.6 : For the following data, find (a) the level of income Y_t for any period and (b) the warranted rate of growth.

$I_t = 3.88\,(Y_t - Y_{t-1})$ $S_t = 0.18Y_t$ $Y_0 = 7000$

a. In equilibrium,

$$3.88\,(Y_t - Y_{t-1}) = 0.18Y_t$$

$$3.88\,Y_t - 3.88Y_{t-1} = 0.18Y_t$$

$$3.88\,Y_t - 0.18Y_t = 3.88Y_{t-1}$$

$$3.7Y_t = 3.88Y_{t-1}$$

Dividing through by 3.7 and then using $Y_t = (Y_0 - \frac{a}{1-b})\,b^t + \frac{a}{1-b}$,

$Y_t = 1.049Y_{t-1} = (7000 - 0)\ (1.049)^t + 0 = 7000(1.049)^t$

b. $G_w = \frac{s}{a - s}$

$G_w = \frac{0.18}{3.88 - 0.18} = 0.048$

2.4.3 Lagged Income Determination Model

The lagged income determination model explains how national income adjusts over time due to a lag in consumption behavior. In reality, people base their current consumption not only on their current income but also on past (lagged) income, leading to dynamic behaviour. The model assumes that (i) the economy is closed (no exports/imports), (ii) only two components of aggregate demand: Consumption (C) and Investment (I), (iii) consumption depends on previous period's income and (iv) investment is autonomous and constant over time.

In the simple income determination model, there were no lags. Now, assume that consumption is a function of the previous period's income, so that

$C_t = C_0 + cY_{t-1}$ $\qquad\qquad$ $Y_t = C_t + I_t$

Where $I_t = I_0$. Thus, $Y_t = C_0 + cY_{t-1} + I_0$

Rearranging the terms to conform with $Y_t = bY_{t-1} + a$

$Y_t = cY_{t-1} + C_0 + I_0$

Where b = c and a = $C_0 + I_0$. Substituting these values in $Y_t = (Y_0 - \frac{a}{1-b})\ b^t + \frac{a}{1-b}$, since the marginal propensity to consume c cannot equal 1, and assuming $Y_t = Y_0$ at t = 0

$Y_t = (Y_0 - \frac{C_0 + I_0}{1 - c})\ (c)^t + \frac{C_0 + I_0}{1 - c}$

The stability of the time path thus depends on c. Since $0 < MPC < 1$, $|c| < 1$ and the time path will converge. Since $c > 0$, there will be no oscillations. The equilibrium is stable, and as $t \to \infty$, $Y_t \to (C_0 + I_0)/\ (1 - c)$, which is intertemporal equilibrium level of income.

The economic interpretation of the lagged income determination model is that the system converges to equilibrium over time, the speed of convergence depends on the size of c and the lag in consumption leads to gradual adjustment in income.

Example.7 : Given $Y_t = C_t + I_t$, $C_t = 150 + 0.85Y_{t-1}$, $I_t = 120$, and $Y_0 = 3000$. Solving for Y_t,

$Y_t = 150 + 0.85Y_{t-1} + 120 = 0.85Y_{t-1} + 270$

Using $Y_t = (Y_0 - \frac{a}{1-b})\ b^t + \frac{a}{1-b}$

$$Y_t = (3000 - \frac{270}{1-0.85})\,(0.85)^t + \frac{270}{1-0.85} = 1200\,(0.85)^t + 1800$$

With $|0.85| < 1$, the time path converges; with $0.85 > 0$, there is no oscillation. Thus, Y_t is dynamically stable. As $t \to \infty$, the first term on the right hand side goes to zero, and Y_t approaches the intertemporal equilibrium level of income. $270/(1-0.85) = 1800$.

To check this answer, let t = 0 and t =1 in $1200\,(0.85)^t + 1800$. Thus,

$$Y_0 = 1200\,(0.85)^0 + 1800 = 3000$$

$$Y_1 = 1200\,(0.85)^1 + 1800 = 2820$$

Substituting $Y_1 = 2820$ for Y_t and $Y_0 = 3000$ for Y_{t-1} in $0.85Y_{t-1} + 270$

$$2820 - 0.85(3000) = 270$$

$$2820 - 2550 = 270$$

Example 8 : Given the data below, (a) find the time path of national income Y_t; (b) Check your answer, using t = 0 and t = 1; and (c) Comment on the stability of the time path

$C_t = 90 + 0.8Y_{t-1}$, $I_t = 50$, and $Y_0 = 1200$. Solving for Y_t,

a. In equilibrium $Y_t = C_t + I_t$. Thus,

$Y_t = 90 + 0.8Y_{t-1} + 50 \quad = \quad 0.8Y_{t-1} + 140$

Using $\quad Y_t = (Y_0 - \frac{a}{1-b})\,b^t + \frac{a}{1-b}$

$$Y_t = (1200 - \frac{140}{1-0.8})\,(0.8)^t + \frac{140}{1-0.8} = 500\,(0.8)^t + 700$$

b. $Y_0 = 1200$; $Y_1 = 1100$. Substituting in $Y_t = 90 + 0.8Y_{t-1} + 50 \quad = \quad 0.8Y_{t-1} + 140$

$1100 = 0.8(1200) + 140 \qquad 1100 = 1100$

c. With b = 0.8, b > 0 and $|b| < 1$. The time path Y_t is nonoscillating and convergent. Y_t converges to the equilibrium level of income 700.

Example 9 : Given the data below, (a) find the time path of national income Y_t; (b) Check your answer, using t = 0 and t = 1; and (c) Comment on the stability of the time path

$C_t = 250 + 0.4Y_t$, $I_t = 2.5(Y_t - Y_{t-1})$, and $Y_0 = 8000$. Solving for Y_t,

a. In equilibrium $Y_t = C_t + I_t$. Thus,

$$Y_t = 250 + 0.4Y_t + 2.5(Y_t - Y_{t-1}) = -1.9Y_{t-1} = -2.5Y_{t-1} + 250$$

Dividing through by -1.9 and then using

$$Y_t = (Y_0 - \frac{a}{1-b})\,b^t + \frac{a}{1-b}$$

$$Y_t = 1.316\,Y_{t-1} - 132\,(8000 - \frac{132}{1-1.316})\,(1.316)^t + \frac{132}{1-1.316} = 7582\,(1.316)^t + 418$$

b. $Y_0 = 8000$; $Y_1 = 10396$. Substituting in the initial equation

$$10396 = 250 + 0.4(10396) + 2.5(10396\text{-}8000) = 10396$$

c. With b = 1.316, the time path Y_t explodes but does not oscillate.

Summarised Overview

Difference equations serve as a powerful mathematical tool for analyzing these dynamics in discrete time periods, particularly when studying investment, consumption, and policy impacts. Central to this analysis is the concept of dynamic stability, which assesses whether an economic system returns to equilibrium after a disturbance. The Cobweb model represents this by depicting how supply and demand adjust over successive periods when production decisions are based on previous prices, often resulting in convergent, divergent, or oscillatory price movements. Similarly, the Lagged Income Determination model demonstrates how income evolves in response to past investment, highlighting the effects of time lags in economic adjustment. These models, grounded in difference equations, provide critical insights into economic fluctuations, stability, and the intertemporal effects of policy and expectations.

Assignments

1. Why are difference equations particularly suitable for modeling discrete-time economic processes?
2. How does the Cobweb model use lagged price information to determine supply decisions?
3. Explain how the Lagged Income Determination model illustrates the relationship between investment and income over time.
4. How can insights from dynamic models like the Cobweb and Harrod's model inform economic policy?
5. Let the model $C_t = 0.6Y_{t-1}$, $I_t = 0.4(Y_{t-1} - Y_{t-2})$, $Y_t = C_t + I_t$.

   a. Compute Y_t for t = 0, 1, 2, 3, given $Y_{t-1} = 400$, $Y_0 = 500$

   b. Does the system appear to be stable?

6. Compare and contrast the Cobweb Model and the Lagged Income Determination Model. Discuss their assumptions, mathematical formulations, and implications for economic policy.

7. A simple linear difference equation is given by

   $Y_t = 0.6Y_{t-1} + 80$

   a. Solve the equation assuming $Y_0 = 200$

   b. Analyse the long run behaviour of the system

   c. Is the system stable?

8. Consider a market where demand and supply are given by:

   $D_t = 250 - 4P_{t,}\ S_t = 20 + 2P_{t-1}$

   a. Derive the price difference equation

   b. Find the equilibrium price

   c. Analyse whether the system is dynamically stable

9. Given $C_t = 0.5Y_{t-1}$, $I_t = 0.3(Y_{t-1} - Y_{t-2})$, $Y_t = C_t + I_t$.

   a. Using initial conditions $Y_{-1} = 400$, $Y_0 = 500$ compute Y_1, Y_2 and Y_3

   b. Comment on the system's stability based on the results.

10. What role do lags play in dynamic economic modeling? Illustrate how time lags in consumption and investment decisions affect macroeconomic equilibrium using appropriate examples.

References

1. Edward T Dowling, Schaum's Outline Series (2001), Introduction to Mathematical Economics, Third Edition, McGRAW-HILL

2. Sreenath Baruah(2012): Basic Mathematics and its applications in Economics, Macmillan India Ltd.

3. Mehta and Madnani (2022): Mathematics for Economists, Sultan Chand & Sons, New Delhi

4. Monga G S (2014), Mathematics and Statistics for Economics, Vikas Publishing House Pvt Ltd

5. Yamane, Taro. (2012). Mathematics for Economists: An Elementary Survey. New Delhi: Prentice Hall of India.

Suggested Readings

1. Chiang, A.C. (2008), Fundamental Methods of Mathematical Economics, McGraw Hill, New York.

2. Y.P. Agarwal: Statistical Methods: Concepts, Application and Computation, Sterling Publishers 1986

3. Hooda R.P: Statistics for Business and Economics , Mac Million, New Delhi

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

3

BLOCK

Statistics

1

UNIT

Introduction to Statistics and Data Collection

Learning Outcomes

After completing this unit, the learner will be able to:

- discuss the importance of statistics in data-based decision-making within organisations
- explain the main uses of statistics in economics
- describe the primary and secondary methods of data collection and their applications
- compare the census and sampling methods used in statistical studies

Background

Our lives revolve around numbers and data, often without us realising it. For example, Syam's basketball coach tracks players' points, rebounds, and assists to decide team positions for the next season. Syam's mother uses statistics at her pharmaceutical job to analyse how medicines work and their side effects, helping doctors make better decisions. Weather forecasts rely on data like temperature and rainfall to predict upcoming weather, while companies use statistics to understand customers by tracking likes, clicks, and purchases to improve marketing. Even a simple grocery trip involves data collection, like noting item prices, quantities, and preferences to plan budgets and track spending habits. In every aspect of life such as sports, medicine, weather, business, or shopping - statistics help us analyse data and make better decisions, showing how powerful and useful it is in understanding the world around us.

Keywords

Statistics, Census, Sampling, Primary Data, Secondary Data

3.1.1 Statistics

Statistics is the science of collecting, organising, analysing, and interpreting numerical data. The term "statistics" is derived from the German word *Statistik* and the Latin word *Status*, both meaning state or government. It was first used by Gottfried Achenwall, a German mathematician known as the "Father of Statistics." Statistics helps people understand complex information and make informed decisions in various fields, from ancient times to the present day. For instance, Egyptian kings collected data as early as 3050 BC to construct the pyramids. Over time, the use of statistics expanded beyond governance to many other areas of life, such as science, economics, business, health, and politics. It now plays a key role in understanding data and supporting decision-making, making it an essential tool for human advancement.

The meaning of statistics can be defined in two ways: in the plural form and in the singular form. In the plural form, it refers to numerical data such as income, population, or prices. In the singular form, it refers to the science of collecting, organising, analysing, and interpreting data, and finally, drawing conclusions from it. Both these aspects of statistics are important. If the data collected is poor or inaccurate, then the decisions made based on that data will also be incorrect. Similarly, even if the data is good, but the methods used to analyse it are weak, we will not be able to extract meaningful or useful information. Therefore, for effective decision-making, both reliable data and sound statistical methods are essential.

According to A.L. Bowley, (i) statistics is the science of counting, (ii) it may rightly be called the science of averages, and (iii) it is the science of measuring the social organism as a whole in all its manifestations. In the words of Boddington, statistics is the science of estimates and probabilities. Another statistician, W.I. King, defined statistics in a broader context: the science of statistics is the method of judging collective, natural, or social phenomena from the results obtained through analysis, enumeration, or the collection of estimates. According to Seligman, statistics is a science that deals with the methods of collecting, classifying, presenting, comparing, and interpreting numerical data gathered to shed light on any area of enquiry. From the above definitions, we can identify some key characteristics of statistics as follows:

a. Statistics are Aggregates of Facts: A single number is not statistics. For example, the national income of a country for one year is not statistics. But if we have data for two or more years, it becomes statistics.

b. Statistics are Affected by Many Factors: Statistical data are influenced by various factors. For example, the sale of a product depends on price, quality, competition, consumer income, etc.

c. Statistics must be Reasonably Accurate: If the data are wrong or misleading, the conclusions will also be wrong. So, accuracy is very important in statistics.

d. Statistics are Collected Systematically: Data must be collected in a proper and organised way. If collected carelessly, the data will not be reliable.

e. Statistics are Collected for a Specific Purpose: Data should be gathered with a clear objective in mind. Without a purpose, the data may not be useful.

f. Statistics must be Comparable and Related: The data collected should be related to each other and suitable for comparison over time or across places.

3.1.2 Statistics in Economics

Statistics play a crucial role in economics, providing the tools and methods needed to analyse, interpret, and present data for informed decision-making and policy formulation. Economists use statistical techniques to study relationships between variables, test hypotheses, forecast future trends, and evaluate the impact of policies. For instance, statistics are employed to calculate key economic indicators such as GDP, inflation rates, and unemployment levels, which help gauge the health of an economy. Regression analysis, a statistical method, is commonly used to determine how factors like interest rates or government spending influence economic growth. An example of a statistical application in economics is analysing the effect of education on income levels. By collecting data on individuals' years of schooling and their corresponding earnings, economists can use statistical models to estimate the degree to which additional education increases income, accounting for other variables such as experience and location. This type of analysis provides valuable insights for policymakers aiming to design effective education and labour market policies, underscoring the indispensable role of statistics in understanding and addressing economic challenges.

3.1.3 Methods of Collecting Data

Data collection is the process of gathering information or facts systematically to understand, analyse, and make decisions about a specific issue or phenomenon. Effective data collection helps in making sound decisions and in framing appropriate economic policies and welfare programs. It also supports the prediction of future trends like inflation, unemployment, or growth rates. Moreover, data collection is the backbone of both academic and professional research, as well as future planning. Mainly, there are two methods used for collecting data, which are

1. Primary Data Collection
2. Secondary Data Collection

3.1.3.1 Primary Data Collection

Primary data collection involves the collection of original data directly from the source or through direct interaction with the respondents. This method allows researchers to obtain firsthand information specifically tailored to their research objectives. There are various methods for primary data collection. They are as follows:

- **Surveys and Questionnaires:** Researchers design structured questionnaires or surveys to collect data from individuals or groups. These can be conducted through face-to-face interviews, telephone calls, mail, or online platforms.

- **Interviews:** Interviews involve direct interaction between the researcher and the respondent. They can be conducted in person, over the phone, or through video conferencing. Interviews can be structured (with predefined questions), semi-structured (allowing flexibility), or unstructured (more conversational).

- **Observations:** Researchers observe and record behaviours, actions, or events in their natural setting. This method is useful for gathering data on human behaviour, interactions, or phenomena without direct intervention.

- **Experiments:** Experimental studies involve the manipulation of variables to observe their impact on the outcome. Researchers control the conditions and collect data to draw conclusions about cause-and-effect relationships.

- **Focus Groups:** Focus groups bring together a small group of individuals who discuss specific topics in a moderated setting. This method helps in understanding opinions, perceptions, and experiences shared by the participants.

3.1.3.2 Secondary Data Collection

Secondary data collection involves using existing data collected by someone else for a purpose different from the original intent. Researchers analyse and interpret this data to extract relevant information. Secondary data can be obtained from various sources, including:

- **Published Sources**: Researchers refer to books, academic journals, magazines, newspapers, government reports, and other published materials that contain relevant data.

- **Online Databases:** Numerous online databases provide access to a wide range of secondary data, such as research articles, statistical information, economic data, and social surveys.

- **Government and Institutional Records:** Government agencies, research institutions, and organizations often maintain databases or records that can be used for research purposes.

- **Publicly Available Data:** Data shared by individuals, organisations, or communities on public platforms, websites, or social media can be accessed and utilized for research.

- **Past Research Studies:** Previous research studies and their findings can serve as valuable secondary data sources. Researchers can review and analyse the data to gain insights or build upon existing knowledge.

3.1.4 Methods of Statistical Investigation

Statistical investigation refers to a systematic process of collecting, organising, analysing, and interpreting data to draw meaningful conclusions. There are two popular methods of statistical investigation, viz: Census Method and Sample Method.

3.1.4.1 Census Method

You have a small classroom with only 10 students. The teacher wants to know the average height of the students in the class. One way to find this would be to measure the height of every single student in the classroom and then calculate the average. In this approach, where the teacher collects data from each and every unit (student) in the population (the classroom), it is known as the Census Method. By measuring the height of all 10 students, the teacher is conducting a complete enumeration or a census of the population. This method ensures that no individual is left out, and the resulting average height calculated from the data will be an accurate representation of the entire classroom.

When the investigator collects information from all the units and elements, it is called the census method. Under this, information about every unit of the aggregate is collected. This method is also known as the Complete Enumeration Method. It is a complete enumeration of the entire population, leaving no unit uncounted or unobserved. This method is often employed when the population size is relatively small or when the study demands a comprehensive understanding of the entire population. Population census is an example of a census method. Under the population census, information is obtained about every household and every person. This information is expensive. The results derived from this method are authentic and reliable. This method is costly because labour and time involved are high.

Merits of a Census Method

- **Intensive Study:** Under census investigation, data must be obtained from each and every unit of the population. Furthermore, it enables the statistician to study more than one aspect of all items in the population. For example, the Indian Government conducts a census investigation once every 10 years. Authorities collect data regarding population size, males and females, education levels, sources of income, religion, etc.

- **Reliable Data:** The data that a statistician collects through a census investigation is more reliable, representative, and accurate. This is because, in a census, the statistician observes every item personally.

- **Suitable Choice:** It is a suitable choice in situations where the different items in the population are not homogeneous.

- **The Basis of Various Surveys:** Data from a census investigation is used as a basis for various surveys.

Demerits of the Census Method

1. **High Costs:** A census investigation is a very costly method because every item in the population must be observed. This requires substantial resources and is usually adopted by government organisations for detailed data collection, such as population censuses or agricultural surveys.

2. **Time-Consuming:** The census method is time-intensive, requiring significant amounts of time and labour to collect and process the data.

3. **Possibility of Errors:** There are several potential sources of error in census investigations, including non-responses, measurement issues, inaccurate definitions of statistical units, and even personal biases of the investigators.

4. **Unsuitability for Large Populations:** The census method becomes impractical for large populations. It is more suitable for smaller populations where data collection can be more manageable.

5. **Difficulty in Data Management:** Due to the large volume of data collected, managing, organizing, and analysing the data can be complex and challenging. This often requires advanced systems and technology.

6. **Inflexibility:** Once data is collected in a census, it is difficult to make changes or updates. If new information becomes relevant or if an error is discovered after the data collection, rectifying it can be cumbersome.

3.1.4.2 Sampling Method

A researcher engaged in a study may find that examining every unit in a population is expensive and time-consuming. Therefore, the researcher uses a sample to estimate the overall characteristics of the population. This approach is known as the Sampling Method. The Sampling Method refers to a statistical technique used to study and draw conclusions about a large population by selecting and analysing a smaller, representative group from it. For instance, suppose a college wants to find out the average study time of its students. The college has 2,000 students, but it is not practical to ask every student. So, the teacher decides to take a random sample of 100 students and asks them how many hours they study each day. After collecting and analysing the data from these 100 students, the teacher finds that the average study time is 3.5 hours per day. Based on this sample, the teacher can reasonably estimate that the average study time for all students in the college is approximately 3.5 hours daily.

Types of Sampling:

1. Random Sampling

Random Sampling refers to a method where every member of the population has an equal chance of being selected. Suppose a college wants to study the experiences of students in their online classes. For this purpose, a teacher writes the names of all class divisions on slips of paper, places them in a box, and randomly picks 10 class divisions. Then, a few divisions from the college are asked about their online class experiences.

2. Systematic Sampling

Systematic Sampling refers to selecting items at regular intervals from a list, starting from a random point. For instance, a farm manager wants to examine the health of crops in a large field. They decide to take samples by selecting every 10th plant along a row, starting from a random point in the field. This random starting point helps eliminate bias in choosing where to begin, ensuring that the sample is more representative of the entire field.

3. Stratified Sampling

Stratified Sampling refers to dividing the population into groups or strata based on a specific characteristic and then taking samples from each group. For instance, a government agency wants to understand crop yields in a region with different types of farms (e.g., small, medium, and large). They divide the farms into these categories or strata and select a sample from each group to ensure that all types of farms are represented.

4. Cluster Sampling

Cluster Sampling refers to dividing the population into clusters, then randomly selecting some clusters and surveying all members within them. For example, a researcher wants to study irrigation practices in a rural region with multiple villages. Instead of surveying all the villages, they randomly select a few villages (clusters) and survey all the farmers within those selected villages.

5. Convenience Sampling

Convenience Sampling means that samples are chosen based on ease of access or availability. For instance, a local agriculture extension officer is conducting a survey on pesticide usage and decides to interview farmers who are attending a nearby agricultural seminar, as they are easily accessible.

6. Quota Sampling

Quota Sampling means that a fixed number of subjects is selected from different groups, but not using random methods. For instance, a researcher wants to study the types of fertilizers used by different farmers in a region. They decide to interview 50 farmers who use organic fertilizers and 50 who use chemical fertilizers. The farmers are not selected randomly, but the researcher ensures that the quota for each group is met.

Advantages of Sampling Method

1. Saves Time

Since only a small part of the population is studied, data can be collected and analysed faster. This makes it ideal when decisions need to be made quickly. It is especially useful in urgent situations like product testing or quick surveys.

2. Less Expensive

Sampling reduces the cost of materials, travel, and manpower. Organisations with limited budgets prefer sampling as it is more affordable than studying the entire population.

3. Practical

Sometimes it is not possible to study the entire population due to its size or inaccessibility. For example, testing every bulb produced in a factory would destroy the bulbs - sampling avoids this.

4. Easy to Check

With fewer data points, it is easier to recheck and verify information. Errors can be identified and corrected more quickly, improving the reliability of findings.

5. Suitable for Large Populations

Sampling is useful when the population is too big to study completely. For example, in a national health survey, researchers study only selected individuals from different regions.

6. Flexibility

Sampling allows the use of different techniques suited to various kinds of research. Researchers can choose from methods like random, stratified, or cluster sampling depending on the situation.

Disadvantages of Sampling Method

1. Sampling Errors

Since a sample is only a part of the population, the results may not match the entire group. There is always a risk that important details from the rest of the population may be missed.

2. Bias

If the sample is chosen unfairly or without proper method, it may give wrong results. For instance, selecting only students from one department to represent the whole college may cause bias.

3. Less Accurate

Estimates from a sample are not always as accurate as full population data. Accuracy depends on the sample size and how well it represents the population.

4. Limited Information

A sample may not provide deep insights into smaller groups or rare events. Important patterns in subgroups might be missed if those groups are not included in the sample.

5. Non-response

Some people selected in the sample may not respond to questions or surveys. This creates gaps in the data and may lead to incorrect results if too many do not participate.

6. Hard to Generalize

If the sample is too small or poorly chosen, the results cannot be applied to the whole population. This limits the usefulness of the study in drawing overall conclusions.

7. Quality of Sampling Frame

A sampling frame is the list or source used to select samples. If the list is incomplete or outdated, some groups may be left out by mistake.

8. Still Can Be Costly

Though cheaper than a census, sampling still needs trained staff and planning. For high-quality results, larger and well-planned samples are required, which may increase costs.

Summarised Overview

Statistics is the science of collecting, organising, analysing, and interpreting numerical data. The word statistics was derived from the Latin and German words meaning “state and the word was first systematically used by Gottfried Achenwall. Statistics has two meanings: in the plural form, it refers to numerical data, while in the singular form, it refers to the methods used to analyse such data. According to thinkers like Bowley, Boddington, and Seligman, statistics is not only about numbers but also about interpretation and estimation, with key characteristics including accuracy, comparability, and systematic collection. In economics, statistics plays a vital role by providing tools to study economic relationships, forecast trends, and evaluate policies using techniques like regression analysis. Data collection is central to statistical analysis. The data can be collected from primary and secondary sources. Primary data is gathered firsthand using methods like surveys, interviews, observations, and experiments, while secondary data is collected from existing sources such as books, reports, databases, and past studies. Statistical investigation can follow two main methods-census and sampling. The census method involves collecting data from every unit in the population, ensuring accuracy, but it is costly and time-consuming. In contrast, the sampling method involves studying a smaller, representative group, which is more practical and economical, though it may introduce errors or bias. Types of sampling include random, systematic, stratified, cluster, convenience, and quota sampling. Each has specific applications depending on the nature of the population. While sampling offers advantages like time and cost savings, ease of use, and flexibility and it also has drawbacks such as sampling errors, bias, and limited generalisability. Hence, both methods serve different needs and must be chosen wisely based on the study’s objective and population size.

Assignments

1. Define statistics and explain its importance in the field of economics with suitable examples.

2. Discuss the various definitions of statistics given by different scholars. What are the key characteristics of statistics?

3. Differentiate between primary data and secondary data. Explain the methods used to collect each type with examples.

4. What is the census method of data collection? Explain its merits and demerits.

5. What do you understand by the sampling method? Describe any four types of sampling techniques with examples.

6. Compare and contrast the census method and the sampling method.

References

1. Gujarathi, D. Sangeetha, N. (2007). *Basic Econometrics* (4th ed.) New Delhi: McGraw-Hill

2. Koutsoyianis, A. (1977). *Theory of Econometrics* (2nd ed.). London. The Macmillan Press Ltd

Suggested Readings

1. Anderson, D., D. Sweeney and T. Williams (2013): "*Statistics for Business and Economics*", Cengage Learning: New Delhi.

2. Goon, A.M., Gupta, and Das Gupta B (2002): *Fundamentals of Statistics* (Vol I), World Press.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

2 UNIT Measures of Central Tendency

Learning Outcomes

After completing this unit, learners will be able to:

- understand the concept of central tendency
- get an idea about how to compute and explain the concept of dispersion
- identify key differences and similarities between central tendency and dispersion measures.
- explore the concept of skewness and its impact on data distribution and interpretation

Background

Statistics is the systematic study of data, encompassing processes such as data collection, classification, tabulation, analysis, interpretation, and presentation to support decision-making across diverse fields. Through these techniques, information is gathered and summarised, enabling meaningful conclusions to be drawn about population characteristics. Once collected, data is classified and tabulated, paving the way for effective analysis and interpretation.

In statistics, a measure of central tendency is a single value that represents an entire dataset, capturing its central focus. For example, a student's report card might show an aggregate percentage to summarise their overall performance across various subjects in a single value. This provides a general sense of performance, but it doesn't reveal the full story. To understand the variability in the data, how much individual scores deviate from the average, we need measures of dispersion. While central tendency measures like the mean, median, and mode offer a central or representative value, measures of dispersion provide insights into the spread or variability of data points around this central value. Together, these measures give a more complete picture of both the typical value and the extent of variation in a dataset.

Keywords

Arithmetic mean, Median. Mode, Range, Standard Deviation, Variance

3.2.1 Measures of Central Tendency

Measures of central tendency are statistical tools used to identify the center or average of a data set. The central value, or average, is commonly referenced in everyday situations, for example, when discussing the average weight of a child or the average income of individuals. However, in statistics, the term average has a more precise meaning: it refers to the value within a distribution that best represents the entire group. Positioned between the highest and lowest values, this central value serves as a useful summary, making it a widely accepted measure of central tendency. According to John I. Griffin, "An average may be thought of as a measure of central value." Similarly, A. L. Bowley stated that "Averages are statistical constants which enable us to comprehend in a single effort the significance of the whole. A good measure of central tendency holds certain important properties. These are:

1. It should be based on all observations in the dataset,
2. It should be rigidly defined,
3. It should be easy to compute and understand.,
4. It should be least affected by extreme values,
5. It should show minimum fluctuation from sample to sample drawn from the same population,
6. It should be capable of further algebraic treatment.

We know that the measures of central tendency are statistical tools used to identify a single representative value that summarises an entire dataset. To find the average or central value, various measures of central tendency are used. They are:

1. Arithmetic Mean
2. Geometric mean
3. Harmonic mean,
4. Median
5. Mode
6. Quartile, Octile, Deciles and Percentiles.

3.2.1.1 Arithmetic Mean

In everyday language, what many refer to as an "average" is formally known to statisticians as the arithmetic mean. The arithmetic mean is calculated by summing all the values in a dataset and dividing the total by the number of values. It represents a single, central value that reflects the overall distribution of the data. This makes it the most commonly used and widely understood measure of central tendency.

A. Properties of the Arithmetic Mean

The arithmetic mean of a distribution has the following mathematical properties.

1. The sum of item deviations of all items from the arithmetic mean in a data set is always zero.

   i.e., $\sum(x - \bar{x}) = 0$

2. The sum of squares of the deviations of the items in a dataset is the minimum when the deviation is taken from the arithmetic mean.

   i.e., $\sum(x - a)^2$ is least when $a = \bar{x}$

3. If the mean of n observations, $x_1, x_2, \dots . . x_n$ is $\bar{x}$, then the mean of the observation,

4. $(x_1 \pm a), (x_2 \pm a), \dots \dots, (x_n \pm a)\ is\ (\bar{x} \pm a).$

5. If each observation is multiplied by p, $p \neq 0$, then the mean of the new observation is $p\bar{x}$.

B. Merits And Demerits of The Arithmetic Mean

The Arithmetic mean is one of the most widely used measures of central tendency. It is calculated by summing all the values in a dataset and dividing by the total number of observations.

The merits and demerits of this measure are outlined below:

1. It has a rigid definition.
2. It is simple to comprehend and compute.
3. Based upon all the observations
4. It is least affected by sampling fluctuations
5. It can be subjected to further mathematical analysis

Demerits Of the Arithmetic Mean are:

1. Extreme value has a significant impact for calculating the mean
2. It cannot be determined by inspection
3. It cannot be used to measure qualitative characteristics like honesty, beauty, cleverness, and so on.
4. It is impossible to calculate for open-ended classes.
5. It is not suitable for averaging ratios and percentages.

C. Computation of the Arithmetic Mean

The method of computation depends on the nature of the data. Let us discuss the various method of computation of data.

1. Individual Series

If x is the variable that takes the values $x_1, x_2, \ldots.. x_n$ over N items, then the mean of x, denoted by $\bar{x}$ is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{N} = \frac{\sum x}{N}$$

where $\sum$ denote the summation is over all values of X.

2. Discrete series

There are two ways to calculate the mean for a discrete series. They are direct method and short cut method.

i. Direct method

Mean $\bar{x} = \frac{\Sigma f \times x}{N}$

where, f-frequency

N- Total number of observations.

ii. Short cut method

This method uses an assumed mean and deviations taken from that assumed mean to determine the arithmetic mean. This method is also known as the deviation method. The assumed mean is chosen as some number midway between the largest and smallest of the observations.

Arithmetic mean $\bar{x} = A + \frac{\Sigma f \times d}{N}$

where, A – Assumed mean

d - Deviation of the observations from the assumed mean $ie,\ (x - A)$

f - frequency

N - Number of observations.

3. Continuous series

i. Direct method

Arithmetic Mean $\bar{x} = \frac{\Sigma f \times x'}{N}$

where, x'- mid-point of various class

f-frequency

N - Total number of observations.

ii. Shortcut method

Arithmetic mean $\bar{x} = A + \frac{\Sigma f \times d'}{N}$

where, A-Assumed mean

d' - Deviations of midpoints from the assumed mean $ie,\ (x' - A)$

f -frequency

N - Number of observations.

iii. Step deviation method

This method is used when class intervals are equal

Arithmetic mean $\bar{x} = A + \frac{\Sigma f \times d''}{N} \times c$

where, A-Assumed mean

d'' - Deviations of mid-points from the assumed mean $ie, d'' = \frac{x' - A}{c}$

c - class interval

f - frequency

N - Number of observations.

Illustration 3.2.1

The marks obtained by ten students in a class test are 45, 40, 37, 18, 17, 35, 10, 28, 36, and 47. Find the average mark in the class.

Solution

The average score of the class test is

$$\overline{X} = \frac{\sum x}{n}$$

$$= \frac{45+40+37+18+17+35+10+28+36+47}{10}$$

$$= \frac{313}{10}$$

=31.3

Therefore, the average mark of the class is 31.3.

Illustration 3.2.2

From the following data of marks obtained by 60 students of a class, calculate the mean mark.

Marks	20	30	40	50	60	70
No of students	8	12	20	10	6	4

Solution

Direct method

Marks	**No of students** **(f)**	***f* x *x***
20	8	160
30	12	360
40	20	800
50	10	500
60	6	360
70	4	280
Total	N= 60	2460

$$\bar{x} = \frac{\Sigma(f \times x)}{N}$$

$$= \frac{2460}{60}$$

$$= 41$$

Therefore Mean Mark is 41

Illustration 3.2.3

The heights in inches of 70 employees in an office are given below. Find the mean height of an employee.

Height (in inches)	60	62	63	65	67	68
No of employees	5	10	12	18	15	10

Solution

Calculation of Arithmetic Mean (Assumed Average =63)

Height (x)	No of employees (f)	dx (x-63)	fdx
60	5	-3	-15
62	10	-2	20
63	12	0	0
65	18	2	36
67	15	4	60
68	10	5	50
Total	**70**		**111**

$$\bar{x} = A + \frac{\sum fdx}{N}$$

$$= 63 + \frac{111}{70}$$

$= 63 + 1.586$

$= 64.586$ (in inches)

Illustration 3.2.4

The weekly observation on cost of living index in a certain City for a particular year is given below. Compute the average weekly cost of living index.

Cost of living Index	140-150	150-160	160-170	170-180	180-190	190-200
No. of weeks	8	12	25	12	8	4

Solution: We shall use the deviation method by taking A =165 the assumed mean.

Class	Class Mark (Y_i)	Frequency (f_i)	d(X-A)	fd
140-150	145	8	-20	-160
150-160	155	12	-10	-120
160-170	165	25	0	0

170-180	175	12	10	120
180-190	185	8	20	160
190-200	195	4	30	120
TOTAL		**N = 69**		**∑fx = 120**

$$\bar{x} = A + \frac{\Sigma(f \times d)}{N}$$

$$\bar{x} = 165 + \frac{120}{69}$$

$$= 165 + 1.739$$

$$= 166.739$$

Illustration 3.2.5

Find the Arithmetic Mean from the following data.

Age of Members	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Persons Dying	20	40	59	78	120	130	135	145

Solution

Age	**f**	**Mid(X)**	**d'' = (X-35)/10**	**f d''**
0-10	20	5	-3	-60
10-20	40	15	-2	-80
20-30	59	25	-1	-59
30-40	78	35	0	0
40-50	120	45	1	120
50-60	130	55	2	260
60-70	135	65	3	405
70-80	145	75	4	580
	727			1166

$$\bar{x} = A + \frac{\Sigma(f \times d)}{N}$$

$$= 35 + \frac{1166}{727} \times 10$$

$$= 35 + 16.044$$

$$= 51.04$$

Illustration 3.2.6

Given below is the following frequency distribution of weights of 60 oranges.

Weight (in gram)	65-84	85-104	105-124	125-144	145-164	165-184	185-204
Frequency	9	10	17	10	5	4	5

Find out how much an orange weighs on average.

Solution

Assumed Mean A = 134.5

Weight	**Mid Value (x)**	**f**	**dx**	**f × dx**
65-84	74.5	9	-60	-540
85-104	94.5	10	-40	-400
105-124	114.5	17	-20	-340
125-144	134.5	10	0	0
145-164	154.5	5	20	100
165-184	174.5	4	40	160
185-204	194.5	5	60	300
Total		60		-720

$$\bar{x} = A + \frac{\Sigma(f \times dx)}{N}$$

$$= 134.5 - \frac{720}{60} = 134.5\text{-}12 = 122.5$$

Step deviation method

This technique is an extension on the short cut technique. The method can be used to quickly determine the mean when the figures of deviations appear to be large and divisible by a common factor. The figures of deviations are reduced using this method by dividing them all by a common factor and multiplying the total of the deviations' products by the same common factor. The formula for this method is as follows:

$$\bar{x} = A + \frac{\Sigma(f \times dx')}{N} x\, C$$

Where,

$$dx' = \frac{dx}{c}$$

c –Common factor by which each of the deviation is divided.

$$dx = x - A$$

Illustration 3.2.7

The following table shows the results of an examination for 80 students. Calculate the mean.

Marks:	0-10	10-20	20-30	30-40	40-50	50-60
No of students:	8	10	22	25	10	5

Solution

Calculation of Arithmetic mean (Assumed mean= 25)

Marks	**Mid Value (x)**	**f**	**dx**	$\frac{dx}{10}$ **dx′**	**fdx′**
0-10	5	8	-20	-2	-16
10-20	15	10	-10	-1	-10
20-30	25	22	0	0	0
30-40	35	25	10	1	25
40-50	45	10	20	2	20
50-60	55	5	30	3	15
Total		80			34

$$\bar{x} = 25 + \frac{34}{80} \text{ x } 10$$

$$= 25 + \frac{340}{80}$$

$$= 25 + 4.25$$

$$= 29.25$$

Illustration 3.2.8

The frequency distribution given below gives the cost of production of sugar cane in different holdings. Find the mean.

Cost in Rs.	0-20	20-40	40-60	60-80	80-100
Frequency	41	51	64	38	7

Solution

Assumed mean = 50

Cost in Rs.	**frequency**	**Mid.x**	**dx = x-50**	**dx′= dx/20**	**f*dx′**
0-20	41	10	-40	-2	-82
20-40	51	30	-20	-1	-51
40-60	64	50	0	0	0
60-80	38	70	20	1	38
80-100	7	90	40	2	14
Total	201				-81

$$\bar{x} = A + \frac{\Sigma(f \times dx')}{N} x\ C$$

$$= 50 - \frac{81}{201} \times 20$$

$$= 50 - 8.06$$

$$= 41.9$$

D. Combined Arithmetic Mean

When a series is made up of two or more component series, the mean of the entire series can be easily defined in terms of the component series means. If $\bar{x}_1$ and $\bar{x}_2$ are the means of two groups of n_1 and n_2 observations, the mean of the combined group of n_1 and n_2 observation is

$$\bar{x}_{12} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

$\bar{x}_{12}$ = Combined Mean

n_1 = Number of items in the first series

n_2 = Number of items in the second series

$\bar{x}_1$ = Arithmetic mean of first series

$\bar{x}_2$ = Arithmetic mean of second series

Illustration 3.2.9

The average score obtained by a group of 80 students on an examination was found to be 40. Another group of 150 students received a mean score of 45 on the same exam. Calculate the average score for both groups together.

Solution

$n_1 = 80, \bar{x}_1 = 40$

$n_2 = 150, \bar{x}_2 = 45$

We know that the combined mean is given by

$$= \bar{x}_{12} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

$$= \frac{80 \text{ x } 40+150 \text{ x } 45}{80+150}$$

$$= \frac{3200+6750}{230}$$

$$= \frac{9950}{230}$$

= 43.26

Illustration 3.2.10

The average weight of 150 students in a class is 60 kilogrammes. The average weight of the boys in the class is 70 kg, while the average weight of the girls is 55 kg. Find the number of boys and the number of girls in the class.

Solution

Combined mean, $\bar{x}_{12}$= 60 kgs

Mean weight of boys, $\bar{x}_1$ = 70kgs

Mean weight of girls, $\bar{x}_2$= 55kgs

Total number of students = 150

Let there be 'x' be boys in the class. Therefore, the number of girls in the class is 150-x

$$\bar{x}_{12} = \frac{n_1\,\bar{x}_1 + n_2\,\bar{x}_2}{n_1 + n_2}$$

$$60 = \frac{70x + 55(150-x)}{150}$$

9000 = 70x + 8250-55x

15x = 750

$$x = \frac{750}{15}$$

=50

So, the number of boys in the class is 50 and the number of girls in the class is 150-50 = 100.

Correction in Mean

The process for correction in mean is as follows

i. Find the sum of the values

ii. Subtract incorrect value from the total

iii. To the total, add the correct value

iv. Divide the total by number of items.

Illustration 3.2.11

The mean wage of 120 factory workers was found to be ₹17,000. It was then discovered that an amount of ₹18750 wage was misread as ₹17850. Find the right mean.

Solution

$$\bar{x} = \frac{\sum x}{n}$$

$$17000 = \frac{\sum x}{120}$$

$\sum x$ = 17000 x 120 =2040000

Incorrect $\sum x$ = 2040000

Correct $\sum x$ = incorrect $\sum x$- incorrect item + correct item

Correct $\sum x$ = 2040000 – 17850 +18750

= 2040900

Correct Mean = $\frac{\text{Correct } \sum X}{n}$

$= \frac{2040900}{120}$

=17007.5

Illustration 3.2.12

140 students were studying in a school. Their mean mark was 45. Later on, it was discovered that the marks of two students were misread as 18 and 13, instead of 58 and 54. Find the correct mean

Solution

$\overline{x} = \frac{\sum x}{n}$

$45 = \frac{\sum x}{140}$

$\sum$x = 140 x 45 =6300

Incorrect $\sum$x = 6300

Correct $\sum$x = incorrect $\sum$x- incorrect item + correct item

Correct $\sum$x = 6300 – (18+13) + (58+54)

= 6300 – 31 + 112

= 6381

Correct Mean = $\frac{\text{Correct } \sum X}{n}$

$= \frac{6381}{140}$

= 45.58

D. Geometric Mean

The geometric mean is the nth root of the product of n observations in the data set. The fundamental formula for its computation is:

GM = $(x_1 \times x_2 \times x_3 \times \ldots . \times x_n)^{\frac{1}{n}}$

When there are more than two items, the computation is simplified by using a logarithm. The formula above can be expressed as:

GM = Antilog of $\frac{1}{N}$ (log x_1+log x_2+⋯+log x_n)

When there is a multiplicative relationship between the data or when the data is compounded, the geometric mean performs well. When the data is nonlinear and particularly when a log transformation is used, geometric mean is used.

Advantages of geometric mean

The following are the advantages of geometric mean:

i. The Geometric Mean is significant since it gives less weight to extreme numbers. As a result, the impact of extremely small and extremely high values is minimised.

ii. It can be further algebraically treated.

iii. It can be used to calculate average, percentage changes, ratios, etc.

iv. It is based on all the observation of the series.

v. It can be used to measure relative changes.

vi. The best average in the construction of index numbers is the geometric mean.

vii. It is rigidly defined.

Limitations of geometric mean

The geometric mean has the following limitations:

i. The geometric mean will not be calculated if some of the observations are negative.

ii. It is tough for a layman to comprehend.

iii. If one or more observations are zero, the geometric mean computation is meaningless because the observation's product is always 0, and hence the geometric mean is zero.

iv. It can sometimes give a value that is not in the series.

1. For individual series

GM = antilog of $\frac{\Sigma \log x}{N}$

Illustration 3.2.13

Wages of 10 workers in a factory given below

85, 15, 500, 70, 75, 250, 45, 8, 36, 40

Find geometric mean.

Solution

x	log x
85	1.9294
15	1.1761
500	2.6990

70	1.8451
75	1.8751
250	2.3979
45	1.6532
8	0.9031
36	1.5563
40	1.6021
	$\sum \log x = 17.6373$

The value of log x is determined from logarithm table

GM = antilog of $\frac{\Sigma \log x}{N}$

= antilog of $\frac{17.6373}{10}$

= antilog of 1.76373

= 58.04

2. For discrete series

GM= antilog of $\frac{\Sigma f \times \log x}{N}$

Where,

x – Value of the variable

N – Number of items

f – Frequency

Illustration 3.2.14

Following are the price and demand for Apple per Kilogram. Find geometric mean of the following.

Price:	130	350	260	250	175	150
Demand:	12	2	4	5	8	10

Solution

x	**f**	**log x**	**f log x**
130	12	2.1139	25.3668

350	2	2.5441	5.0882
260	4	2.4150	9.66
250	5	2.3979	11.9895
175	8	2.2430	17.944
150	10	2.1761	21.761
	41		$\sum$ f log x =91.8095

GM= antilog of $\frac{\Sigma f \log x}{N}$

= antilog of $\frac{91.8095}{41}$

= antilog of 2.2392

=173.5

3. For continuous series

GM = antilog of $\frac{\Sigma f \log x}{N}$

Where,

x – Mid value

Illustration 3.2.15

From the following data calculate Geometric mean

Income (in 000):	0-10	10-20	20-30	30-40
No of families:	5	8	3	4

Solution

Income	**Mid value (x)**	**f**	**log x**	**f log x**
0-10	5	5	0.6990	3.4950
10-20	15	8	1.1761	9.4088
20-30	25	3	1.3979	4.1937
30-40	35	4	1.5441	6.1764
		20		23.2739

GM = antilog of $\frac{\Sigma f \log x}{N}$

= antilog of $\frac{23.2739}{20}$

= antilog of 1.1637

= 14.58

Income is shown in thousands. Therefore, the Geometric Mean = 14580

Illustration 3.2.16

From the following data calculate Geometric mean

Marks:	0-10	10-20	20-30	30-40	40-50
No of students:	5	7	15	25	8

Solution

Marks	**f**	**midx**	**log x**	**f log x**
0-10	5	5	0.6990	3.4950
10-20	7	15	1.1761	9.4088
20-30	15	25	1.3979	4.1937
30-40	25	35	1.5441	6.1764
40-50	8	45	1.6532	13.2256
Total	60			84.5243

GM = antilog of $\frac{\Sigma f \log x}{N}$

= antilog of $\frac{84.5243}{60}$

= antilog of 1.4087

= 25.62

E. Harmonic Mean

The Harmonic mean of a number of observations is the reciprocal of the arithmetic mean of the reciprocal of the given observations.

Harmonic mean can be defined as “the reciprocal of the arithmetic average of the reciprocal of the value of a variable”.

$$\therefore \text{H.M} = \frac{1}{\frac{1}{N}\left(\frac{1}{x_1}+\frac{1}{x_2}+\dots+\frac{1}{x_n}\right)} = \frac{\mathbf{N}}{\sum\frac{1}{\mathbf{x}}}$$

Where,

H.M – Harmonic Mean

N – Number of items

x – Value of the variable

When we wish to average units like speed, rates, and ratios, we use the harmonic mean.

Advantages of harmonic mean

The benefits of Harmonic Mean are as follows:

i. It gives the smallest item the most weight.

ii. It is quite beneficial for averaging certain ratios and rates because it measures relative changes.

iii. It is based on all the observations

iv. It is rigidly defined

v. It is possible to calculate it even if a series contains any negative numbers

Limitations of harmonic mean

The following are the limitations of Harmonic Mean

i. It is very difficult to calculate

ii. It does not accurately reflect the statistical series.

iii. It is tough for a layman to comprehend.

iv. It is impossible to calculate if any of the items are zero

v. This is merely a summary figure; the actual item in the series may not be shown

vi. It has a very limited algebraic treatment.

1. For individual observation

$$\text{H.M} = \frac{\mathbf{N}}{\sum\frac{1}{\mathbf{x}}}$$

Illustration 3.2.17

The speeds of five buses in a city are given below.

Speed (Km/hr):	15	18	20	22	17

Find the average speed

Solution

X	$\frac{1}{x}$
15	0.0666
18	0.0555
20	0.05
22	0.04545
17	0.05882
$\Sigma\frac{1}{x}$ = 0.27637	

$$H.M = \frac{N}{\Sigma\frac{1}{x}}$$

$$H.M = \frac{5}{0.27637}$$

$$= 18.09$$

Average speed = 18.09 Km/hr

2. For discrete observation

$$H.M = \frac{N}{\Sigma f\frac{1}{x}}$$

Illustration 3.2.18

Calculate the harmonic mean of the scores on the English class test, as shown below.

Marks:	11	12	13	14	15
No of students:	8	7	4	5	2

Solution

x	f	$\frac{1}{x}$	$f\frac{1}{x}$
11	8	0.0909	0.7272
12	7	0.0833	0.5831

13	4	0.0769	0.3076
14	5	0.0714	0.357
15	2	0.0666	0.1332
	26		2.1081

$$\text{H.M} = \frac{N}{\Sigma f \frac{1}{x}}$$

$$= \frac{26}{2.1081}$$

$$=12.33$$

3. For continuous observations

$$\text{H.M} = \frac{N}{\Sigma f \frac{1}{x}}$$

Where,

x – Mid Value of the classes

Illustration 3.2.19

From the following data, calculate Harmonic mean

Class:	0-10	10-20	20-30	30-40
Frequency:	2	3	4	2

Solution

Class	**Mid Value (x)**	**f**	$\frac{1}{x}$	$f\frac{1}{x}$
0-10	5	2	0.2	0.4
10-20	15	3	0.066	0.2
20-30	25	4	0.04	0.16
30-40	35	2	0.0285	0.0571
		11		0.8171

$$H.M = \frac{N}{\sum f \frac{1}{x}}$$

$$= \frac{11}{0.8171}$$

$$=13.462$$

3.2.1.2 Median

The median is the middle value of a dataset when the numbers are arranged in either ascending or descending order. Alternatively, the median may be defined as that value of the variable which divides the group into two equal parts-one part consisting of all values greater than the median and the other consisting of all values less than the median. The way the median is calculated depends on the number of data points:

- If the number of observations is odd, the middle value is the median.
- If the number of observations is even, the median is the average of the two middle values.

A. Advantages of the Median

i. It is rigidly defined

ii. It's simple to calculate the median.

iii. Extreme values or outliers have no effect on it.

iv. It is possible to calculate it for open-ended classes.

v. When dealing with qualitative data where no numerical measurements are provided but it is possible to rank the objects in some order, the median is the only average that can be employed.

vi. The absolute sum of the individual values' deviations from the median is always the minimum.

B. Disadvantages of the Median

i. Median is not suitable for further mathematical treatment

ii. In the case of ungrouped data with an even number of observations, the median cannot be estimated precisely. The arithmetical average of the two middle elements is the median in this case.

iii. It sometimes generates a value that is not seen anywhere else in the series.

iv. To calculate the median, the data must be arranged in ascending or descending order.

v. The median is less stable than the mean, especially in small samples.

vi. The median, as a positional average, does not take into account every single item in the distribution.

C. Computation of median

1. For individual series

Steps

i. Sort the data into ascending or descending order.

ii. Use the formula.

Median = $\left(\frac{n+1}{2}\right)^{th}$ item

Illustration 3.2.20

The marks obtained by a student in five examinations are given below.

Marks:	35	37	25	28	40

Find the median mark.

Solution

Arrange the data in ascending order

25	28	35	37	40

Apply the formula

Median = $\left(\frac{n+1}{2}\right)^{th}$ item

= $\left(\frac{5+1}{2}\right)^{th}$ item

=3rd item

The 3rd item in the series is 35.

∴ Median mark is 35

Illustration 3.2.21

The following table shows the income of six families. Find their median income.

Income:	10000	12000	11000	20000	15000	17000

Solution

Arrange the data in ascending order

10000	11000	12000	15000	17000	20000

Apply the formula

Median = $\left(\frac{n+1}{2}\right)^{th}$ item

= $\left(\frac{6+1}{2}\right)^{th}$ item

=3.5th item

However, there is not a single item in the series with a position of 3.5. As a result, we use the median as the average of the third and fourth elements in the series.

Median = Mean of 3rd and 4th item

$= \frac{12000+15000}{2}$

$= \frac{27000}{2}$

=13500

Median income of the family is 13500.

2. For discrete series

Steps

Arrange the data in ascending or descending order

Calculate cumulative frequency (cf)

Determine $\frac{N+1}{2}$

Where N is the total frequency

Median is the value for the $(\frac{N+1}{2})^{th}$ item of the data

Illustration 3.2.22

The daily wage of 115 employees is shown in the table below. Find out what the median wage is.

Wage:	500	600	700	800	900	1000	1100	1200
No of employees:	8	14	15	18	20	15	14	11

Solution

Wages	f	cf
500	8	8
600	14	22
700	15	37

800	18	55
900	20	75
1000	15	90
1100	14	104
1200	11	115
	N=115	

Median = $\frac{N+1}{2}$

$= \frac{115+1}{2}$

$= \frac{116}{2}$

= 58th item

∴ Median is the value in the data which comes in the 58th position, which is the value of the item having cumulative frequency 58. Since cumulative frequency of 58 comes under the cumulative frequency 75, median is the value in the data that comes in the 75th position,

∴ Median = 900

3. For continuous series

Steps

i. Convert inclusive classes to the exclusive class (if any)

ii. Calculate the cumulative frequencies (cf)

iii. Calculate $\frac{N}{2}$, where N is the total frequency

iv. Identify the class having cumulative frequency $\frac{N}{2}$

v. Find median by using this formula;

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

Where,

l - Lower limit of the median class

m - Cumulative frequency of the class preceding the median class.

f - Frequency of the median class

c - Class interval of the median class

Illustration 3.2.23

The following table shows the household income of 80 families.

Income (in'000):	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No of household:	12	10	13	19	13	8	5

Find median income

Solution

The cumulative distribution table is

Class	**F**	**cf**
0-10	12	12
10-20	10	22
20-30	13	35
30-40	19	54
40-50	13	67
50-60	8	75
60-70	5	80
	N = 80	

$$\frac{N}{2} = \frac{80}{2} = 40$$

The class having cumulative frequency 40 is 30-40

∴ Median class is 30-40

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \text{ X C}$$

$$= 30 + \frac{40 - 35}{19} \text{ x } 10$$

$$= 30 + \frac{50}{19}$$

$$= 30 + 2.63$$

$$= 32.631$$

Illustration 3.2.24

The table below shows the distribution of marks obtained in English by 265 students in the Science and Commerce streams.

Mark:	0-9	10-19	20-29	30-39	40-49	50-59	60-69
No of students:	15	40	50	60	45	40	15

Find Median

Solution

Here the classes are of the inclusive type. Before computing the median, the inclusive class should be converted into an exclusive class to get the actual class limit.

Marks	**Actual class**	**f**	**cf**
0-9	-0.5 - 9.5	15	15
10-19	9.5 – 19.5	40	55
20-29	19.5 – 29.5	50	105
30-39	29.5 - 39.5	60	165
40-49	39.5 – 49.5	45	210
50-59	49.5 – 59.5	40	250
60-69	59.5 – 69.5	15	265
		N = 265	

$\frac{N}{2} = \frac{265}{2} = 132.5$

The class having cumulative frequency 132.5 is 29.5 – 39.5

∴ Median class is 29.5 - 39.5

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

$$= 29.5 + \frac{132.5 - 105}{60} \times 10$$

$$= 29.5 + \frac{275}{60}$$

$$= 29.5 + 4.583$$

$$= 34.083$$

Illustration 3.2.25

Calculate the median mark from the following frequency table giving the distribution of marks of 80 students.

Mark:	0-9	10-19	20-29	30-39	40-49
No of students:	3	10	28	36	3

Solution

Here the classes are of the inclusive type. Before computing the median, the inclusive class should be converted into an exclusive class to get the actual class limit.

Marks	Actual class	f	cf
0-9	-0.5 – 9.5	3	3
10-19	9.5 – 19.5	10	13
20-29	19.5 – 29.5	28	41
30-39	29.5 – 39.5	36	77
40-49	39.5 – 49.5	3	80
		N = 80	

$$\frac{N}{2} = \frac{80}{2} = 40$$

The class having cumulative frequency 40 is 19.5 – 29.5

∴ Median class is 19.5 - 29.5

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \text{ x c}$$

$$= 19.5 + \frac{(40-13)}{28} \text{ x } 10$$

$$= 19.5 + \frac{270}{28}$$

$$= 19.5 + 9.64$$

$$= 29.14$$

3.2.3.3 Mode

The mode of a data set is defined as the value that appears the most frequently in the data. It is the data observation with the highest frequency. Mode means norm or fashion. Mode is also known as the business average or fashionable average.

Advantages of Mode

i. It is possible to calculate it graphically.

ii. It determines which value in a series is the most representative.

iii. It is unaffected by the series' extreme values.

iv. It has a lot of applications in the sphere of business and commerce.

v. It is simple and clear to compute and comprehend.

vi. It is not required to know the values of all the items in a series to calculate mode.

vii. The position of mode is likewise not a problem with open ended classes.

viii. The only average that works with categorical data is the mode.

Disadvantages of mode

i. Because it is not rigidly defined, it may have different results in some instances.

ii. Further algebraic treatment is not possible.

iii. It is not based on all evidence.

iv. It is ill-defined, indefinite, and ambiguous.

v. It can only be calculated from series with unequal class intervals if they are equalised.

vi. It is influenced by sampling fluctuations.

Computation of mode

i. For individual series

The mode in individual observations is the most recurring value in a series.

Illustrations 3.2.26

Eleven people aged 18, 17, 19, 18, 17, 18, 21, 22, 18, 23, and 21 years old took part in a cricket match. Determine the mode of the data.

Solution

Here the observation 18 appears 4 times, 17 and 21 appears 2 times and all others are appeared in a single time. So, the value which appears a maximum number of times is 18.

∴ Mode = 18 years

ii. For Discrete series

Observation with highest frequency is considered as the mode in the discrete series.

Illustration 3.2.27

The age distributions in the following graphs illustrate the age of employees in various departments. Determine the mode.

Age:	21	22	23	24	25	26
No of employees:	5	15	50	30	21	17

Solution

The age 23 has the highest frequency. Therefore, 23 is the mode.

Grouping Table and Analysis Table

The item with the highest frequency is referred to as a mode. However, if the maximum frequency is repeating or if the maximum frequency occurs at the beginning or end of the distribution or if there are irregularity in the distribution it may be impossible to find the mode simply by looking at it. In rare circumstances, the frequency concentration may be more concentrated around a frequency that is lower than the highest frequency. A grouping table and an analysis table should be developed to determine the correct modal value in such circumstances.

Steps for calculation

i. Construct a six-column grouping table.

ii. In column (1), record the frequency in relation to the item.

iii. The frequencies in column (2) are arranged in twos, starting at the top. Their totals are calculated, and the highest total is highlighted.

iv. The frequencies are grouped in twos again in column (3), leaving the first frequencies. The highest total is once again noted.

v. The frequencies in column (4) are arranged in threes, starting at the top. Their totals are calculated, and the highest total is highlighted.

vi. The frequencies are grouped in threes again in column (5), leaving the initial frequency. Their totals are calculated, and the highest total is highlighted.

vii. The frequencies are grouped in threes again in column (6), with the first and second frequencies leaving. After totalling the frequencies, the highest total is identified and highlighted again.

viii. Create an analysis table to find the modal value or modal class that the largest frequencies cluster around for the longest periods of time. Place the column number on the left-hand side of the table and the item sizes on the right-hand side. Mark 'X' in the relevant box corresponding to the values they represent to input the values against which the highest frequencies are found. The mode is the set of values with the most 'X' marks against them.

Illustration 3.2.28

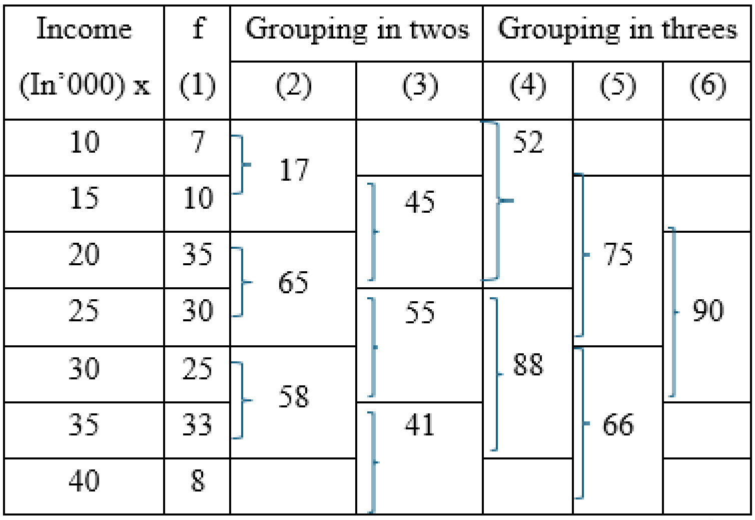
The following table shows the monthly income of 130 families. Calculate the mode value.

Income (in'000):	10	15	20	25	30	35	40
No of Families:	7	10	35	30	25	33	8

Solutions

Since there are irregularity in the distribution, we must construct the grouping table and analysis table because determining the modal value is tough by examination.

(a)Grouping table



Income (In'000) x	f (1)	Grouping in twos (2)	(3)	Grouping in threes (4)	(5)	(6)
10	7	17		52		
15	10		45		75	
20	35	65				90
25	30		55	88		
30	25	58			66	
35	33		41			
40	8					

In column 1 the highest frequency is 35 corresponds to the 20. So, we put X mark in 20. In column 2 the highest frequency is 65 corresponds to 35 and 30. So we put X mark in 20 and 25. In column 3 the highest frequency is 55 corresponds to 30 and 25. So we put X mark in 25 and 30. In column 4 the highest frequency is 70 corresponds to 30,25 and 15. So we put X mark in 25,30 and 35. In column 5 the highest frequency is 75 corresponds to 10,35 and 30. So we put X mark in 15, 20 and 25. In column 6 the highest frequency is 90 corresponds to 35, 30 and 25. So we put X mark in 20,25 and 30.

(b) Analysis table

Variable **F column**	**10**	**15**	**20**	**25**	**30**	**35**	**40**
1			X				
2			X	X			
3				X	X		
4				X	X	X	
5		X	X	X			
6			X	X	X		
Total	-	1	4	5	3	1	-

The greatest total (5) is noted to be against the value of 25. As a result, the modal mark is 25.

Mode =25

iii. For continuous series

Steps

Locate the modal class having highest frequency or by preparing analysis table.

Apply this formula

$$\text{Mode} = l + \frac{(f_1 - f_0)}{2f_1 - f_0 - f_2} \text{ X C}$$

where, l – Lower limit of the modal class.

f_1 – Frequency of the modal class.

f_0 – Frequency of the preceding class to the modal class.

f_2 – frequency of the succeeding class to the modal class.

c – Class interval of the modal class

Illustration 3.2.29

The following table shows the frequency distribution of the marks of 80 students. Find the mode.

Marks:	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No.of students:	3	5	7	10	12	15	14	4	2

Solution

Here 50-60 is the model class. c = 10, f_1=15, f_2=14, f_0 = 12

$$\text{Mode} = l + \frac{(f_1 - f_0)}{2f_1 - f_0 - f_2} \text{ X C}$$

$$= 50 + \frac{(15-12)}{2*15-12-14} \text{ x } 10$$

$$= 50 + \frac{30}{4}$$

$$= 57.5$$

Illustration 3.2.30

The following table shows the monthly income of 130 families. Calculate the mode value.

Income (in'000):	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No of Families:	4	8	18	30	20	10	30	3	2

Solution

Here the maximum frequency 30 is repeating, we use grouping table and an analysis table **Grouping Table**

Income (In'000)	f (1)	Grouping in twos		Grouping in threes		
		(2)	(3)	(4)	(5)	(6)
0-10	4	12		30		
10-20	8		26		56	
20-30	18	48				68
30-40	30		50	60		
40-50	20	30			60	
50-60	10		40			43
60-70	30	33		35		
70-80	3		5			
80-90	2					

(b) Analysis table

Variable	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
F column								
1				X			X	
2			X	X				
3				X	X			
4				X	X	X		
5					X	X	X	
6			X	X	X			
Total	-	-	2	5	4	2	2	-

The modal class is identified as 30-40. The following formula can be used to calculate mode.

$$\text{Mode} = l + \frac{(f_1 - f_0)}{2f_1 - f_0 - f_2} \text{ X C}$$

l =30

$f_1 = 30$

$f_0 = 18$

$f_2 = 20$

$c = 10$

$$= 30 + \frac{(30 - 18)}{(2 \text{ x } 30) - 18 - 20} \text{ x } 10$$

$$= 30 + \frac{120}{22}$$

$$= 30 + 5.45$$

$$= 35.45$$

3.2.2 Measures of Dispersion

Measures of dispersion are statistical tools that describe the spread, variability, or distribution of data values around a central point (like the mean, median, or mode). While measures of central tendency tell us about the "average" or "middle" of the data, measures of dispersion tell us how much the data varies or is scattered. These measures help in comparing variability in different datasets, understanding consistency in data, and making informed decisions based on the spread of values. The major types of measures of dispersion are:

1. Range,
2. Interquartile Range,
3. Standard Deviation, and
4. Variance.

3.2.2.1 Range

The range is the simplest of all the measures of dispersion. It is defined as the difference between the two extreme observations of the distribution. In other words, range is the difference between the greatest (maximum) and the smallest (minimum) observation of the distribution. It indicates the limit with which the values fall.

Range = L-S

where,

L = Largest value, S = Smallest value

Coefficient of Range = $\frac{L-S}{L+S}$

In case of the grouped frequency distribution (for discrete values) or the continuous frequency distribution, range is defined as the difference between the upper limit of the highest class and the lower limit of the smallest class.

If the average of the two distributions are about the same, a comparison of the range indicate that the distribution with smaller range has less dispersion and the average of that distribution is more representative of the group.

Merits of Range

1. It is simplest to understand and easier to compute as it only involves finding the difference between the maximum and minimum values in a dataset.
2. It provides an intuitive understanding of the spread in data
3. Its simplicity makes it a quick and accessible measure of dispersion.
4. Calculating the range requires only the knowledge of the two extreme values, making it applicable even in situations where detailed data may not be readily available

Demerits of Range

1. Since it is based on extreme values it may be changed either of the extreme values happens to drop out.
2. It does not consider the size of the dataset.
3. It does not take into account the frequencies of the distribution.
4. The range can change drastically with the addition or removal of just one data point.
5. Range cannot be found for open end distributions.

Uses of Range

Range is useful in the following situations,

1. In statistical quality control range can be used as a measure of variation.
2. Range is used to describe the difference between a commodity's highest and lowest price. It is the most widely used measure of variability in our daily lives.
3. For weather forecasts, the meteorological department uses a range.
4. Range can be applied in areas where the data have small variations

Illustration 3.2.31

Below are the prices of 1 kg of sugar for the first six months. Find Range and Coefficient of Range.

Month	January	February	March	April	May	June
Price/kg	125	110	160	130	165	160

Solution

Range = L-S

L =165, S = 110

Range = 165-110

= 55

Coefficient of Range = $\frac{L-S}{L+S}$

$= \frac{165-110}{165+110}$

$= \frac{55}{275}$

= 0.2

For discrete series

Illustration 3.2.32

From the following data relating to the monthly income of 60 people, determine the range and coefficient of range.

Income:	210	240	290	360	440	510	500	350	290
No of person:	5	10	15	7	3	10	2	3	5

Solution

Range = L-S

= 510 – 210

= 300

Coefficient of Range = $\frac{510-210}{510+210}$

$= \frac{300}{720}$

= 0.417

For continuous series

Illustration: 3.2.33

The following table gives the age distribution of a group of 50 individuals.

Age (in years) :	16 – 20	21 – 25	26 – 30	31 – 35
No. of persons :	10	15	17	8

Calculate range and the coefficient of range.

Solution

Since age is a continuous variable, we should first convert the given classes into continuous classes.

Age	x	f
16-20	15.5-20.5	10
21-25	20.5-.25.5	15
26-30	25.5-30.5	17
31-35	30.5.5-35.5	8

Largest value = 35·5; Smallest value = 15·5 ∴ Range = 35·5 – 15·5 = 20 years

$$\text{Coefficient of Range} = \frac{35.5-15.5}{35.5+15.5}$$

$$= \frac{20}{51}$$

$$= 0.39$$

3.2.2.2 Inter Quartile Range

The quartile deviation is a measure of dispersion based on quartiles. Quartiles are the points which divide the array in 4 equal parts. The interquartile range is a measure of dispersion based on the upper quartile Q_3 and lower quartile Q_1. The upper quartile (first quartile) is $\left(\frac{n+1}{4}\right)^{th}$ term and lower quartile (third quartile) is $3 \times \left(\frac{n+1}{4}\right)^{th}$ term of the observations.

Inter quartile range = $Q_3 - Q_1$

Inter quartile range divided by two is the Quartile Deviation. It gives the average amount by which the two quartiles differ from the median. In a symmetric distribution the two quartiles Q_1 and Q_3 are equidistant from the median. Small Quartile Deviation means that the variations among the central items are small and high Quartile Deviation means that the variations among the central items are high.

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

It can be used to compare the degree of variation in different distributions.

Merits of Quartile Deviation

- It is easy to understand and calculate
- It can be calculated for open-ended classes.
- It is unaffected by extreme values.

Demerits of Quartile Deviation

- It is not based on all observation. It ignores 50% of the observations
- It is not capable of further algebraic treatment.
- It is very much affected by samplying fluctuations.
- Computation of inter quartile range

For individual series

Illustration 3.2.34

Compute the inter-quartile range, quartile deviation, and coefficient of quartile deviation from the following data:

X:	20	28	40	12	30	15	50

Solution

Arrange the data in ascending order

X:	12	15	20	28	30	40	50

$n = 7$

Q_1 = value of $\left(\frac{n+1}{4}\right)^{th}$ item

= $\left(\frac{7+1}{4}\right)^{th}$ item

= 2nd item

=15

Q_3 = value of $3\left(\frac{n+1}{4}\right)^{th}$ item

= 3 x 2nd item

= 6th item

= 40

Inter-quartile range = $Q_3 - Q_1$

= 40 - 15

= 25

Quartile Deviation = $\frac{Q_3 - Q_1}{2}$

$= \frac{25}{2}$

$= 12.5$

Coefficient of Quartile Deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{40-15}{40+15}$

$= \frac{25}{55}$

$= 0.46$

Illustration 3.2.35

Compute the inter-quartile range, quartile deviation, and coefficient of quartile deviation from the following data:

X:	14	13	9	7	12	17	8	10	6	15	18	20	21

Solution

Arrange the data in ascending order

X:	6	7	8	9	10	12	13	14	15	17	18	20	21

$n = 13,$

Q_1 = value of $\left(\frac{n+1}{4}\right)^{th}$ item

$= \left(\frac{13+1}{4}\right)^{th}$ item

= 3.5th item

= 3rd item + 0.5 (4th item -3rd item)

= 8 + 0.5 (9-8)

= 8.5

Q_3 = value of $3\left(\frac{n+1}{4}\right)^{th}$ item

= 3 x 3.5th item

= 10.5th item

= 10th item + 0.5 (11th item – 10th item)

= 17 + 05 (18-17)

= 17.5

Inter-quartile range = $Q_3 - Q_1$

$= 17.5 - 8.5$

$= 9$

Quartile Deviation = $\frac{Q_3 - Q_1}{2}$

$= \frac{17.5 - 8.5}{2}$

$= \frac{9}{2}$

$= 4.5$

Coefficient of Quartile Deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{17.5 - 8.5}{17.5 + 8.5}$

$= \frac{9}{26}$

$= 0.346$

For discrete series

Illustration 3.2.36

Below are the heights (in inches) of 43 people.

Height (in inches):	12	20	30	40	50	80
No of persons:	4	7	15	8	7	2

Calculate inter-quartile range, quartile deviation, and coefficient of quartile deviation.

Solution

Height (in inches)	**Frequency**	**Cum.f**
12	4	2
20	7	11
30	15	26
40	8	34
50	7	41
80	2	43
	43	

$n = 43$

Q_1 = Series having cum.f $\left(\frac{43+1}{4}\right)$

= Series having cf 11

= 20

Q_3 = Series having cf 3 $\left(\frac{43+1}{4}\right)$

= Series having cf 33

= 40

Inter-quartile range = $Q_3 - Q_1$

= 40-20

= 20

Quartile Deviation = $\frac{Q_3 - Q_1}{2}$

$= \frac{40-20}{2}$

$= \frac{20}{2}$

= 10

Coefficient of Quartile Deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{40-20}{40+20}$

$= \frac{20}{60}$

$= \frac{1}{3}$

For continuous series

Illustration 3.2.37

Calculate the inter-quartile range, quartile deviation, and coefficient of quartile deviation.

Farm Size (acres)	0-40	41-80	81-120	121-160	161-200	200 above
No of farms	13	17	50	60	55	45

Solution

Farm size	X	f	cf
0-40	-0.5-40.5	13	13
41-80	40.5-80.5	17	30
81-120	80.5-120.5	50	80
121-160	120.5-160.5	60	140
161-200	160.5-200.5	55	195
Above 200	Above 200.5	45	240

Q_1Class $= \left(\frac{n}{4}\right)^{th}$ Class

$= \left(\frac{240}{4}\right)^{th}$ Class

$= 60^{\text{th}}$ Class

Series having cum. frequency 60 is 80.5-120.5 . ∴ Q_1 class is 80.5-120.5

$$Q_1 = l + \frac{(\frac{N}{4} - m_1)}{f_1} \times c_1$$

$$= 80.5 + \frac{(60-30)}{50} \times 40$$

$$= 80.5 + \frac{30 \times 40}{50}$$

$$= 80.5 + \frac{1200}{50}$$

$$= 80.5 + 24$$

$$= 104.5$$

Q_3 Class $= 3\left(\frac{n}{4}\right)^{th}$ Class

$= 3\left(\frac{240}{4}\right)^{th}$ h Class

$= 180^{\text{th}}$ Class

160.5-200.5 is Q_3 class

$$Q_3 = l + \frac{(\frac{3N}{4} - m_1)}{f_1} \times c_1$$

$$= 160.5 + \frac{(180 - 140)}{55} \times 40$$

$$= 160.5 + \frac{40 \times 40}{55}$$

$$= 160.5 + \frac{40 \times 40}{55}$$

$$= 160.5 + 29.09$$

$$= 189.59$$

$Q_1 = 104.5$, $Q_3 = 189.59$

Inter quartile range $= Q_3 - Q_1$

$= 189.59 - 104.5$

$= 85.09$

Quartile Deviation $= \frac{Q_3 - Q_1}{2}$

$= \frac{85.09}{2}$

$= 42.545$

Coefficient of Quartile Deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{189.59 - 104.5}{189.59 + 104.5}$

$= \frac{85.09}{294.09}$

$= 0.289$

Illustration 3.2.38

The following are the height of students in a class. Find the Quartile deviation.

Height (inches)	50-53	53-56	56-59	59-62	62-65	65-68
No of students	2	7	24	27	13	3

Solution

Height	**f**	**c f**
50-53	2	2
53-56	7	9
56-59	24	33
59-62	27	60
62-65	13	73
65-68	3	76

Q_1 Class = $\left(\frac{76}{4}\right)^{th}$ Class

= 19 th Class

56-59 is Q_1 class

$$Q_1 = l + \frac{\left(\frac{N}{4} - m_1\right)}{f_1} \times c_1$$

$$= 56 + \frac{(19-9)}{24} \times 3$$

$$= 56 + \frac{10 \times 3}{24}$$

$$= 56 + 1.25$$

$$= 57.25$$

Q_3 Class = $3\left(\frac{n}{4}\right)^{th}$ Class

= $3\left(\frac{76}{4}\right)^{th}$ h Class

= 57 th Class

59-62 is Q_3 class

$$Q_3 = l + \frac{\left(\frac{3N}{4} - m_1\right)}{f_1} \times c_1$$

$$= 59 + \frac{(57-33)}{27} \times 3$$

$$= 59 + \frac{24 \times 3}{27}$$

$= 59 + \frac{72}{27}$

$= 59 + 2.67$

$= 61.67$

$Q_1 = 57.25, \quad Q_3 = 61.67$

Inter quartile range $= Q_3 - Q_1$

$= 61.67 - 57.25$

$= 4.42$

Quartile Deviation $= \frac{Q_3 - Q_1}{2}$

$= \frac{4.42}{2}$

$= 2.21$

3.2.2.3 Standard Deviation

The most important and widely used measure of dispersion is the Standard deviation. It is the positive square root of the mean of the squares of deviation from the arithmetic mean. It is denoted by the Greek letter σ (sigma). It cannot be negative. The standard deviation concept was introduced by Karl Pearson in 1893. It is the most used methods of dispersion since it is free from some defects of other measures of dispersion.

The square of the Standard deviation σ^2 is termed as variance ans is more often specified than standard deviation. It has the same properties as Standard deviation.

Merits of Standard Deviation

1. It is rigidly defined.
2. It is based on all observation.
3. It is never disregards the plus or minus sign
4. It can be subjected to more mathematical analysis.
5. The changes in sampling have little effect on it.
6. It allows us to compare and contrast two or more series and determine their consistency or stability.
7. It is used in testing of hypothesis

Demerits of standard deviation

1. A layman would find it difficult to comprehend.

2. It is complex to calculate since it incorporates several mathematical models.

3. It cannot be used to compare the dispersion of two or more series of observations with different units of measurement.

3.2.3 Coefficient of Variation

The coefficient of variation is calculated by dividing the standard deviation by the arithmetic mean, which is given as a percentage. It is the most popular way of comparing the consistency or stability of two or more sets of data. The series for which the CV is greater is said to be more variable or less consistent or less stable. On the other hand, the series for which CV is less is said to be less variable or more consistent or more stable.

$$CV = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100$$

$$= \frac{\sigma}{\bar{x}} \times 100$$

Computation of Standard deviation and Coefficient of variation

For individual series

Standard deviation = $\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$

Variance = $\sigma^2 = \frac{\Sigma(x-\bar{x})^2}{n}$

where,

$\bar{x}$ -Actual mean of the observation

n - Total number of items

Illustration 3.2.39

Calculate the standard deviation for the following data

1,3,5,7,4

Solution

$$\bar{x} = \frac{\Sigma x}{n}$$

$$= \frac{20}{5}$$

$$= 4$$

X	(x -4)	$(x-4)^2$
1	-3	9

3	-1	1
5	1	1
7	3	9
4	0	0
		20

Standard deviation = $\sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$

$= \sqrt{\frac{20}{5}}$

$= \sqrt{4}$

$= 2$

CV= $\frac{\text{Standard Deviation}}{\text{Mean}} \times 100$

$= \frac{2}{4} \times 100$

$= 50\%$

For discrete series

Standard deviation = $\sqrt{\frac{\Sigma f \times x^2}{\Sigma f} - \bar{x}^2}$

Where, f – Frequency, $\bar{x} = \frac{\Sigma f \times x}{\Sigma f}$

Illustration 3.2.40

Find the standard deviation for the following data

Size:	6	9	12	15	18
Frequency	7	12	19	10	2

Solution

x	**f**	**fx**	**x^2**	**f x^2**
6	7	42	36	252
9	12	108	81	972

12	19	228	144	2736
15	10	150	225	2250
18	2	36	324	648
Total	**50**	564		6858

$$\bar{x} = \frac{\Sigma f \times x}{\Sigma f}$$

$$= \frac{564}{50}$$

$$= 11.28$$

$$\text{Standard deviation} = \sqrt{\frac{\Sigma f \times x^2}{\Sigma f} - \bar{x}^2}$$

$$= \sqrt{\frac{6858}{50} - 11.28^2}$$

$$= \sqrt{137.16 - 127.24}$$

$$= \sqrt{9.92}$$

$$= 3.15$$

Illustration 3.2.41

Find the standard deviation for the following data. Also find the coefficient of variation.

No. of letters	2	3	4	5	6	7	8	9	10	11
Frequency	9	6	2	2	2	4	3	3	2	3

Solution

x	**f**	**fx**	**x^2**	**fx^2**
2	9	18	4	36
3	6	18	9	54
4	2	8	16	32

5	2	10	25	50
6	2	12	36	72
7	4	28	49	196
8	3	24	64	192
9	3	27	81	243
10	2	20	100	200
11	3	33	121	363
	N = 36	198		1438

$\overline{x} = \frac{\Sigma fx}{N}$

$= \frac{198}{36}$

$= 5.5$

Standard deviation $= \sqrt{\frac{\Sigma f \times x^2}{\Sigma f} - \bar{x}^2}$

$= \sqrt{\frac{1438}{36} - 5.5^2}$

$= \sqrt{39.94 - 30.25}$

$= \sqrt{9.69}$

$= 3.11$

CV $= \frac{\text{Standard Deviation}}{\text{Mean}}$ x 100

$= \frac{3.11}{5.5}$ x100

$= 56.5\%$

Illustration 3.2.42

The score of 2 batsman A and B in 10 innings during a certain match are as under .

A	32	28	47	63	71	39	10	60	96	14
B	19	31	48	53	67	90	10	62	40	80

Who is the better batsman? Who is more consistent?

Solution

In order to decide as to which of the two batsman, A or B, is better player, we should find their average score. The one whose average is higher will be considered as a better batsman.

T determine the consistency we should determine the coefficient of variation. The less this coefficient of variation is more consistent.

$$A's\ \bar{x} = \frac{\Sigma x}{n}$$

$$= \frac{460}{10}$$

$$= 46$$

$$B's\ \bar{x} = \frac{\Sigma x}{n}$$

$$= \frac{500}{10}$$

$$= 50$$

A			**B**		
X	**(x -46)**	**(x - 46)^2**	**X**	**(x -50)**	**(x -50)^2**
32	-14	196	19	-31	961
28	-18	324	31	-19	361
47	1	1	48	-2	4
63	17	289	53	3	9
71	25	625	67	17	289
39	-7	49	90	40	1600
10	-36	1296	10	-40	1600
60	14	196	62	12	144
96	50	2500	40	-10	100
14	-32	1024	80	30	900
		6500			5968

A

Standard deviation = $\sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$

$= \sqrt{\frac{6500}{10}}$

$= \sqrt{650}$

$= 25.5$

CV = $\frac{\text{Standard Deviation}}{\text{Mean}}$ x 100

$= \frac{25.5}{46}$ x100

= 55.4%

B

Standard deviation = $\sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$

$= \sqrt{\frac{5968}{10}}$

$= \sqrt{596.8}$

$= 24.43$

CV = $\frac{\text{Standard Deviation}}{\text{Mean}}$ x 100

$= \frac{24.43}{50}$ x100

= 48.8 %

B is better batsman since his average is 50 as compared to 46 of A.

B is more consistent since the coefficient of variation of B is less than the coefficient of variation of A

iii. For continuous series

Standard deviation = $\sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}}$

OR

Standard deviation = $\sqrt{\frac{\Sigma(fx^2)}{\Sigma f} - (\bar{x})^2}$

Where,

f – Frequency

x – Mid value

Illustration 3.2.43

The marks of 75 students in a class is given below.

Mark:	1-3	3-5	5-7	7-9	9-11	11-13	13-15
No of students:	1	9	25	35	17	10	3

Find the standard deviation and the coefficient of variation of the data.

Solution

Mark	Mid value (x)	F	f x	(x- 8)^2	f(x- 8)^2
1-3	2	1	2	36	36
3-5	4	9	36	16	144
5-7	6	25	150	4	100
7-9	8	35	280	0	0
9-11	10	17	170	4	68
11-13	12	10	120	16	160
13-15	14	3	42	36	108
Total		100	800		616

$$\overline{x} = \frac{\sum fx}{\sum f}$$

$$= \frac{800}{100}$$

$$= 8$$

Standard deviation $= \sqrt{\frac{\sum f(x-\overline{x})^2}{\sum f}}$

$$= \sqrt{\frac{616}{100}}$$

$$= 2.48$$

Illustration 3.2.44

Find the standard deviation and the coefficient of variation of the data.

Wages (Rs)	30-32	32-34	34-36	36-38	38-40	40-42	42-44
No of labours	12	18	16	14	12	8	6

Solution

Mark	F	Mid value (x)	f x	x^2	$f\,x^2$
30-32	12	31	372	961	11532
32-34	18	33	594	1089	19602
34-36	16	35	560	1225	19600
36-38	14	37	518	1369	19166
38-40	12	39	468	1521	18252
40-42	8	41	328	1681	13448
42-44	6	43	258	1849	11094
Total	86		3098		112694

$$\overline{x} = \frac{\sum fx}{\sum f}$$

$$= \frac{3098}{86}$$

$$= 36.02$$

Standard deviation $= \sqrt{\frac{\Sigma(fx^2)}{\Sigma f} - (\overline{x})^2}$

$$= \sqrt{\frac{112694}{86} - 36.02^2}$$

$$= \sqrt{1310 - 1297.44}$$

$$= \sqrt{12.56}$$

$$= 3.54$$

Illustration 3.2.45

Find the standard deviation of the following data.

Age (years)	Less than 10	Less than 20	Less than 30	Less than 40	Less than 50	Less than 60	Less than 70	Less than 80
No of Persons	15	30	53	75	100	110	115	125

Solution

Age	Cum f	f	Mid value (x)	xf	x^2	fx^2
0-10	15	15	5	75	25	375
10-20	30	15	15	225	225	3375
20-30	53	23	25	575	625	14375
30-40	75	22	35	779	1225	26950
40-50	100	25	45	1125	2025	50625
50-60	110	10	55	550	3025	30250
60-70	115	5	65	325	4225	21125
70-80	125	10	75	750	5625	56250
		125		4395		203325

$$\overline{x} = \frac{\sum fx}{\sum f}$$

$$= \frac{4395}{125}$$

$$= 35.16$$

Standard deviation $= \sqrt{\frac{\Sigma(fx^2)}{\Sigma f} - (\overline{x})^2}$

$$= \sqrt{\frac{203325}{125} - 35.16^2}$$

$$= \sqrt{1626.6 - 1236.23}$$

$$= \sqrt{390.37}$$

$$= 19.76$$

Illustration 3.2.46

The standard deviation calculated from a set of 32 observations is 5. If the sum of the observations is 80, what is the sum of the square of the observations?

Solution

Given $n = 32,\ \sigma = 5,\ \sum x = 80$

$$\bar{x} = \frac{\sum x}{n}$$

$$= \frac{80}{32} = 2.5$$

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

$$5 = \sqrt{\frac{\sum x^2}{32} - 2.5^2}$$

Squaring both sides

$$25 = \frac{\sum x^2}{32} - 2.5^2$$

$$25 = \frac{\sum x^2}{32} - 6.25$$

$$\frac{\sum x^2}{32} = 25 + 6.25 = 31.25$$

$$\sum x^2 = 31.25 \times 32 = 1000$$

Illustration 3.2.47

Suppose you are a teacher and you want to analyze the performance of two students.

Rahul	20	22	17	23	28
Manu	10	20	18	12	15

Determine which of the two students, Rahul or Manu, is the most consistent in terms of scoring.

Solution

Rahul's $\bar{x} = \frac{\sum x}{n}$

$= \frac{110}{5}$

$= 22$

Manu's $\bar{x} = \frac{\sum x}{n}$

$= \frac{75}{5}$

$= 15$

Rahul			Manu		
X	**(x -22)**	**(x -22)^2**	**X**	**(x -15)**	**(x -15)^2**
20	-2	4	10	-5	25
22	0	0	20	5	25
17	-5	25	18	3	9
23	1	1	12	-3	9
28	6	36	15	0	0
∑x = 110		66	∑x = 75		**68**

Rahul

Standard deviation = $\sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$

$= \sqrt{\frac{66}{5}}$

$= \sqrt{13.2}$

= 3.63

CV = $\frac{\text{Standard Deviation}}{\text{Mean}}$ x 100

$= \frac{3.63}{22}$ x 100

= 16.5%

Manu

Standard deviation = $\sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$

$= \sqrt{\frac{68}{5}}$

$= \sqrt{13.6}$

= 3.69

CV = $\frac{\text{Standard Deviation}}{\text{Mean}}$ x 100

$= \frac{3.69}{15}$ x 100

= 24.6 %

In comparison to Manu, Rahul is more consistent in his scoring because his coefficient of variation is lower.

Combined standard deviation

The following formula can be used to calculate the combined standard deviation of two or more groups:

$$\sigma_{1.2} = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$$

Where,

$\sigma_{1.2}$ - Combined standard deviation

σ_1 – Standard deviation of the first series

σ_2 - Standard deviation of the second series

d_1 – $(\bar{x}_1 - \bar{x}_{1.2})$

d_2 – $(\bar{x}_2 - \bar{x}_{1.2})$

$\bar{x}_{1.2}$ – Combined mean

$\bar{x}_{1.2}$ - $\frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$

n_1 – Number of items of the first series

n_2 – Number of items of the second series

Illustration 3.2.48

Calculate the combined standard deviation of the two Factories using the given information.

	Factory A	**Factory B**
Mean	63	54
SD	8	7
Number of item	50	40

Solution

$$\sigma_{1.2} = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$$

$$\bar{x}_{1.2} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n+n_2}$$

$$= \frac{(50 \times 63)+(40 \times 54)}{50+40}$$

$$= \frac{3150+2160}{90}$$

$$= \frac{5310}{90}$$

$$= 59$$

$d_1 = (\bar{x}_1 - \bar{x}_{1.2})$

= (63 – 59)

= 4

$d_2 = (\bar{x}_2 - \bar{x}_{1.2})$

= (54 -59)

= -5

$$\sigma_{1.2} = \sqrt{\frac{(50 \times 8^2) + (40 \times 7^2) + (50 \times 4^2) + (40 \times -5^2)}{50 + 40}}$$

$$= \sqrt{\frac{(50 \times 64) + (40 \times 49) + (50 \times 16) + (40 \times 25)}{90}}$$

$$= \sqrt{\frac{3200 + 1960 + 800 + 1000}{90}}$$

$$= \sqrt{\frac{6960}{90}}$$

$$= \sqrt{77.33}$$

= 8.79

Illustration 3.2.49

Analysis of the monthly wages of two hospitals gave the following information.

	Hospital I	Hospital II
No. of staff	550	600
Average wages	60	48.5
Variance	100	144

Obtain the average wage and combined standard deviation of the two hospitals together.

Solution

$$\sigma_{1.2} = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1d_1^2 + n_2d_2^2}{n_1 + n_2}}$$

$$\bar{x}_{1.2} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$$

$$= \frac{(550 \times 60)+(600 \times 48.5)}{550+600}$$

$$= \frac{33000+29100}{1150}$$

$$= \frac{62100}{1150}$$

$= 54$

$d_1 = (\bar{x}_1 - \bar{x}_{1.2})$

$= (60 – 54)$

$= 6$

$d_2 = (\bar{x}_2 - \bar{x}_{1.2})$

$= (48.5 -54)$

$= -5.5$

$$\sigma_{1.2} = \sqrt{\frac{(550 \times 100) + (600 \times 144) + (550 \times 6^2) + (600 \times (-5.5)^2)}{550 + 600}}$$

$$= \sqrt{\frac{55000 + 86400 + 19800 + 18150}{1150}}$$

$$= \sqrt{\frac{179350}{1150}}$$

$= \sqrt{155.96}$

$= 12.49$

Illustration 3.2.50

For a group containing 100 observations the mean $\bar{x} = 8$ and $\sigma = \sqrt{10.5}$. For 50 observations selected from these 100 observations, the mean and standard deviation are 10 and 2 respectively. Find the mean and standard deviation of the other half.

Solution

Given $n_1 + n_2 = 100$, $\bar{x}_{1.2} = 8$, $\sigma_{1.2} = \sqrt{10.5}$, $\bar{x}_1 = 10$, $\sigma_1^2 = 2^2$

$\bar{x}_{1.2} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

$8 = \frac{50\times 10 + 50\times\bar{x}_2}{100}$

$50 \times 10 + 50 \times \bar{x}_2 = 8 \times 100 = 800$

$50 \times \bar{x}_2 = 800 - 500 = 300$

$\bar{x}_2 = \frac{300}{50} = 6$

$d_1 = (\bar{x}_1 - \bar{x}_{1.2})$

$= (10 - 8)$

$= 2$

$d_2 = (\bar{x}_2 - \bar{x}_{1.2})$

$= (6 - 8)$

$= -2$

$$\sigma_{1.2} = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$$

$$\sqrt{10.5} = \sqrt{\frac{(50 \times 4) + (50 \times \sigma_2^2) + (50 \times 4) + (50 \times 4)}{100}}$$

Squaring both sides

$$10.5 = \frac{(50 \times 4) + (50 \times \sigma_2^2) + (50 \times 4) + (50 \times 4)}{100}$$

$(50 \times 4) + (50 \times \sigma_2^2) + (50 \times 4) + (50 \times 4) = 10.5 \times 100$

$(200) + (50 \times \sigma_2^2) + (200) + (200) = 1050$

$(50 \times \sigma_2^2) = 1050 - 200 - 200 - 200 = 1050 - 600 = 450$

$(\sigma_2^2) = \frac{450}{40} = 9$

$\sigma_2 = 3$

Correction in mean and standard deviation

Because one or more observations in a series are inaccurate, the mean and standard deviation computed from the series may be incorrect. The correct values of those observations may be known only after the calculations are over. As a result, we must rectify the mean and standard deviation by taking the correct values of those observations into account.

Illustration 3.2.51

The mean and standard deviation of 11 observations were calculated as 5 and 3.67, respectively. But later, it was identified that one item having a value of 2 was misread as 13. Calculate the correct mean and standard deviation.

Solution

Incorrect $\sum x = \bar{x} \times n$

$= 5 \text{ x } 11$

$= 55$

Correct $\sum x$ = Incorrect $\sum x$ – wrong item + correct item

$= 55–13+2$

$= 44$

Correct $\bar{x} = \frac{44}{11}$

$= 4$

Calculation of the correct Standard Deviation

$$= \sqrt{\frac{\sum x^2}{N} - (\bar{x})^2}$$

$$3.67 = \sqrt{\frac{\sum x^2}{11} - (5)^2}$$

Squaring both sides

$$3.67^2 = \frac{\sum x^2}{11} - 25$$

$$13.4689 + 25 = \frac{\sum x^2}{11}$$

$$38.4689 = \frac{\sum x^2}{11}$$

$\sum x^2 = 38.4689 \text{ x } 11$

Incorrect $\sum x^2 = 423.1579$

Correct $\sum x^2$ = Incorrect $\sum x^2$- square of wrong item + square of correct item.

$= 423.1579 – 13^2 + 2^2$

= 423.1579 -169 + 4

= 258.1579

Correct SD = $\sqrt{\frac{258.1579}{11} - 4^2}$

$= \sqrt{\frac{258.1579}{11} - 16}$

$= \sqrt{23.4689 - 16}$

$= \sqrt{7.4686}$

= 2.73

Illustration 3.2.52

For a group of 200 candidates the mean and standard deviation of scores were found to be 40 and 15 respectively. Later on it was discovered that the score 43 and 35 were wrongly written as 34 and 53 respectively. Find the corrected mean and standard deviation corresponding to the corrected figure,

Solution

Incorrect $\sum x = \overline{x}$ xn

= 40 x 200

= 8000

Correct $\sum x$ = Incorrect $\sum x$ – wrong item + correct item

= 8000 – (34+53) + (43+35)

= 8000 – 87 +78

= 7991

Correct $\overline{x} = \frac{7991}{200}$

= 39.955

Calculation of the correct Standard Deviation

$= \sqrt{\frac{\Sigma x^2}{N} - (\overline{x})^2}$

$15 = \sqrt{\frac{\Sigma x^2}{200} - (40)^2}$

Squaring both sides

$15^2 = \frac{\Sigma x^2}{200} - 1600$

$1600 + 225 = \frac{\sum x^2}{200}$

$1825 = \frac{\sum x^2}{200}$

$\sum x^2 = 1825 \text{ x } 200$

Incorrect $\sum x^2 = 365000$

Correct $\sum x^2$ = Incorrect $\sum x^2$- square of wrong item + square of correct item.

$= 365000 - (34^2 + 53^2) + (43^2 + 35^2)$

$= 365000 - 3965 + 3074$

$= 364109$

Correct SD $= \sqrt{\frac{364109}{200} - 39.955^2}$

$= \sqrt{224.143}$

$= 14.971$

Summarised Overview

Measures of Central Tendency and Dispersion are fundamental statistical tools used to summarize and understand data. Mean, Median, and Mode are the three main measures of central tendency. The mean gives the average value, the median represents the middle value when data is arranged in order, and the mode indicates the most frequently occurring value. These help in identifying the center or typical value of a dataset. On the other hand, measures of dispersion describe the spread or variability within the data. The range is the difference between the highest and lowest values, while the interquartile range (IQR) measures the spread of the middle 50% of data, helping to reduce the effect of outliers. Standard deviation and variance quantify how much data points deviate from the mean, with variance being the square of the standard deviation. Together, these measures provide a comprehensive understanding of both the central value and variability in a dataset.

Assignments

1. The heights in inches of 70 employees in an office are given below. Find the mean height of an employee

Height (in inches)	60	62	63	65	67	68
No of employees	5	10	12	18	15	10

2. Given below is the following frequency distribution of weights of 60 oranges.

Weight (in gram):	65-84	85-104	105-124	125-144	145-164	165-184	185-204
Frequency:	9	10	17	10	5	4	5

Find out how much an orange weighs on average.

3. The mean wage of 120 factory workers was found to be ₹17,000. It was then discovered that an amount of ₹18750 wage was misread as ₹17850. Find the right

4. Find the median wage of the following distribution

Wages (Rs):	20-30	30-40	40-50	50-60	60-70
No of labourers:	3	5	20	10	5

5. The following table represent the income of 122 families. Calculate Median income.

Income:	1000	1500	3000	2000	2500	1800
No of family:	24	26	16	20	6	30

6. Find mode from the following data.

Mark:	20-24	25-29	30-34	35-39	40-44	45-59
No of student:	20	24	32	28	20	26

7. Calculate mode from the following data

Weight in kg:	25	30	35	40	45	50	55	60
No of person:	50	70	80	180	70	30	20	10

8. The average daily wage of all workers in a factory is Rs. 444. If the average daily wages paid to male and female workers are Rs. 480 and Rs. 360 respectively, find the percentage of male and female workers employed by the factory.

9. Calculate Range, Interquartile range and Quartile Deviation from the following data.

X:	4	8	3	9	16	10	14	20	18	15	21

Range = 18, Interquartile range = 10, Quartile Deviation = 5

10. A survey of domestic consumption of electricity in a village gave the following distribution of units consumed.

Units:	Below 100	100-200	200-300	300-400	400-500	500-600	600-700	Above 700
No of Consumers:	20	21	30	46	20	25	16	10

Find Quartile deviation and interquartile range.

Interquartile range = 296, Quartile deviation = 148

11. The arithmetic mean and standard deviation of series of 20 items were calculated by a student as 20 cm. and 5 cm. respectively. But while calculating them an item 13 was misread as 30. Find the correct arithmetic mean and standard deviation.

Mean = 19.15, Standard Deviation = 4.6615

12. The results of ten distinct class tests for two students, Radha and Syama, are shown here.

Radha:	25	50	45	30	70	42	36	48	34	60
Syama:	10	70	50	20	95	55	42	60	48	80

Determine which of the two students, Radha or Syama, is the most consistent in terms of scoring

CV of Radha= 29.37%, CV of Syama= 45.94%, Radha is more consistent

13. Below are the profits earned by 100 sole proprietorship businesses.

Profit in '000:	0-10	10-20	20-30	30-40	40-50	50-60
No of companies:	8	12	20	30	20	10

Calculate the standard deviation and the coefficient of variation of the data.

SD= 13.84, CV = 42.89

14. Obtain the Standard deviation of the following data

Value	90-99	80-89	70-79	60-69	50-59	40-49	30-39
Frequency	2	12	22	20	14	4	1

SD= 12.505

15. Two workers on the Same job show the following results over a long period of time.

	Worker A	Worker B
Mean	30	25
SD	6	4

Which worker appears to be more consistent.

Ans : CV(A)=20, CV(B)=16

References

1. Gujarathi, D. Sangeetha, N. (2007). *Basic Econometrics* (4thed) New Delhi: McGraw Hill

2. Koutsoyianis, A. (1977). *Theory of Econometrics* (2nded). London. The MacMillian Press ltd

Suggested Readings

1. Anderson, D., D. Sweeney and Williams (2013): *"Statistics for Business and Economics",* Cengage Learning: New Delhi.

2. Goon, A.M., Gupta and Das Gupta B (2002): *Fundamentals of Statistics* (Vol I), World Press.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

3

UNIT

Skewness and Kurtosis

Learning Outcomes

After completing this unit, the learner will be able to:

- understand the direction and degree of asymmetry in a dataset.
- evaluate the concentration of data in the tails and the overall peaked-ness of the distribution.
- understand skewness and kurtosis is essential for making informed decisions in statistical analysis.

Background

Statistical measures play a crucial role in unveiling the distinctive characteristics of a frequency distribution, offering insights beyond the basic parameters of central tendency and dispersion. While measures of central tendency, such as the mean, median, and mode, provide a sense of where the bulk of observations cluster within a distribution, and measures of dispersion, like variance and standard deviation, highlight the degree of spread around these central values, they might not capture the full complexity of different distributions.

In certain instances, frequency distributions can exhibit substantial variations in their nature and composition, yet yield identical or similar values for central tendency and dispersion. This is where additional statistical measures become indispensable. Skewness and kurtosis, for instance, looks deeper into the shape and characteristics of a distribution. Skewness gauges the asymmetry of the distribution, indicating whether it is skewed to the left or right, providing valuable information about the tail behaviour. On the other hand, kurtosis measures the degree of peaked-ness or flatness in the distribution, shedding light on the tails' thickness and the overall shape.

we need statistical measures which will reveal clearly the salient features of a frequency distribution. The measures of central tendency tell us about the concentration of the observations about the middle of the distribution and the measures of dispersion give us an idea about the spread or scatter of the observations about some measure of central tendency. We may come across frequency distributions which differ very widely in their nature and composition and yet may have the same central tendency and dispersion.

Keywords

Measure of Skewness, Moments, Measure of kurtosis, Lorentz curve, Gini coefficient

3.3.1 Skewness

We have already introduced two parameters which describe the frequency distribution. They are mean, which locates the distribution and standard deviation which measure the scatter of the items bout the mean. However, they do not reveal the entire story. How symmetrical the distribution is about the central value or how peaky is the distribution are some other features that specify the distribution. They are skewness and kurtosis.

Imagine you are analyzing the household income distribution in a country. In a perfectly symmetrical distribution, most households would cluster around the median income, creating a balanced bell-shaped curve. However, real-life income distributions often exhibit skewness.

If there are fewer extremely high-income households, the income distribution is positively skewed causing the tail on the right side (higher incomes) to be longer or fatter. This could occur in a scenario where a small percentage of the population earns significantly higher incomes, leading to a rightward skew. For instance, a country with a booming tech industry might have a positively skewed income distribution, as a few individuals or households earn exceptionally high salaries or profits.

Conversely, negative skewness in the income distribution would indicate a longer or fatter tail on the left side (lower incomes). This might be observed in situations where a large proportion of the population earns relatively low incomes, but a smaller group has significantly higher earnings. For instance, in an economy with a large low-wage service sector and a small number of high-income executives, the income distribution could be negatively skewed.

Understanding the skewness of the income distribution is valuable for policymakers and researchers. Positive skewness may suggest issues related to income inequality, while negative skewness might indicate challenges associated with widespread low incomes. Skewness, in this context, provides a quantitative measure to describe and interpret the shape of the income distribution, offering insights beyond just average income values.

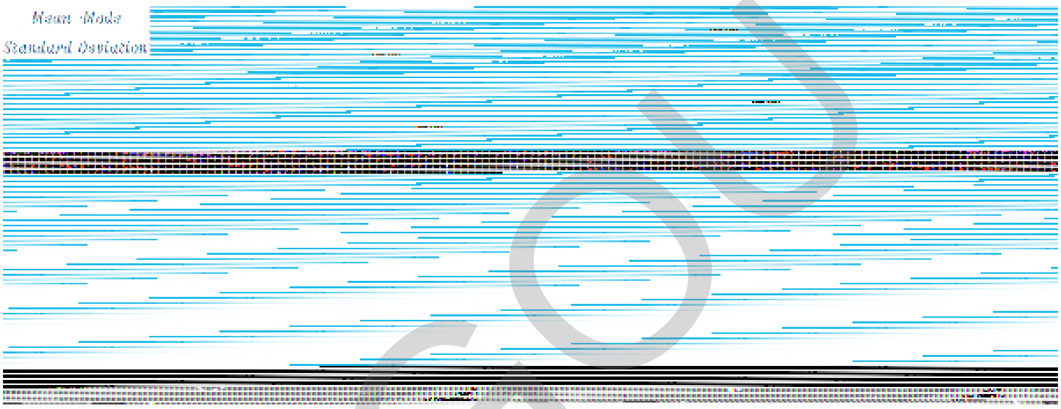
Let us consider mean, mode, and median for three 3 hypothetical distributions, A, B, and C.

For distribution A, Mean = 5, Median = 5, Mode = 5. Here, the value of Mean, Median and Mode are identical. Hence it is known as symmetric distribution.

For distribution B, Mean = 4.06, Median = 4, Mode = 3. In distribution B, Mean is maximum and Mode is least. The excess value on the right hand side. Hence it is known as positively skewed distribution.

For distribution C, Mean = 5.94, Median = 6, Mode = 7. In distribution C, Mean is least and Mode is maximum. The excess value on the left hand side. Hence it is known as negatively skewed distribution.

Let us see the graphical representation based on the distributions.



Distribution B Distribution A Distribution C

3.3.1.1 Measure of Skewness

Measure of Skewness tells us the direction and extend of asymmetry in a series and permit us to compare two or more series.

Karl Pearson's coefficient of Skewness SK= $\frac{Mean-Mode}{Standard\ Deviation}$

If coefficient of Skewness is positive, distribution is positively skewed and if coefficient of Skewness is negative, distribution is negatively skewed.

Illustration – 3.3.1

Compute Karl Pearson's coefficient of Skewness from the following table

Value :	6	12	18	24	30	36	42
Frequency:	4	7	9	18	15	10	5

Solution

Value x	f	xf	x^2	fx^2
6	4	24	36	144
12	7	84	144	1008
18	9	162	324	2916
24	18	432	576	10368
30	15	450	900	13500
36	10	360	1296	12960
42	5	210	1764	8820
Total	68	1722	5040	49560

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{1722}{68} = 25.32$$

$$SD = \sqrt{\frac{49716}{68} - (25.32)^2}$$

$$= \sqrt{731.12 - 641.10}$$

$$= 9.48$$

Mode = 24

Coefficient of Skewness = $\frac{25.32-24}{9.48}$

$= \frac{1.32}{9.48}$

= 0.139

Coefficient of Skewness is positive, the distribution is positively skewed

Illustration – 3.3.2

Compute Karl Pearson's coefficient of Skewness from the following table

Value :	5-7	8-10	11-13	14-16	17-19
Frequency:	14	24	38	20	4

Solution

Value x	X	f	Mid x	xf	x^2	fx^2
5-7	4.5-7.5	14	6	84	36	504
8-10	7.5-10.5	24	9	216	81	1944
11-13	10.5-13.5	38	12	456	144	5472
14-16	13.5-16.5	20	15	300	225	4500
17-19	16.5-19.9	4	18	72	324	1296
Total		100		1128		13716

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{1128}{100} = 11.28$$

$$SD = \sqrt{\frac{13716}{100} - (11.28)^2}$$

$$= \sqrt{137.16 - 127.24}$$

$$= 3.15$$

Highest frequency is 38. So, 10.5-13.5 is the model class

$$Mode = 10.5 + \frac{(38-24)\times 3}{2\times 38 - 24 - 20} = 10.5 + \frac{42}{32} = 11.81$$

Coefficient of Skewness = $\frac{11.28-11.81}{3.15}$

$= - \frac{0.53}{3.15}$

= -0.17

Coefficient of Skewness is negative, the distribution is negatively skewed

Illustration – 3.3.3

Consider the following distributions

	Distribution A	Distribution B
Mean	100	90
Median	90	80
Standard Deviation	10	10

1. Prove that Distribution A is more variable than Distribution B
2. Prove that both Distribution has same measure of skewness.

Solution

1. Coefficient of variation of A = $\frac{\sigma_A}{\bar{x}_A} \times 100 = \frac{10}{100} \times 100 = 10$

Coefficient of variation of B = $\frac{\sigma_B}{\bar{x}_B} \times 100 = \frac{10}{90} \times 100 = 11.11$

Since CV (B) > CV(A), the distribution B is more variable than distribution A

2. Coefficient of Skewness SK= $\frac{Mean-Mode}{Standard\ Deviation}$

The empirical relation is $Mean - Mode = 3(Mean - Median)$

Distribution A $100 - Mode = 3(100 - 90) = 30$

$Mode \quad = 100 - 30 = 70$

SK(A) = $\frac{Mean-Mode}{Standard\ Deviation}$

$= \frac{100-70}{10} = \frac{30}{10} = 3$

Distribution B $90 - Mode = 3(90 - 80) = 30$

$Mode \ = 90 - 30 = 60$

SK = $\frac{Mean-Mode}{Standard\ Deviation}$

$= \frac{90-60}{10} = \frac{30}{10} = 3$

Coefficient of Skewness is same for both distribution

Moments

The term moment is obtained from mechanics where the moment of force is the capacity of a force to turn a pivoted level.

For a frequency distribution, at a distance x from the origine, a force equal to the frequency f associatd with x acts and the moment about the origine is $f \times x$. Taking the whole distribution moment is $\sum f \times x$

If the total moment divided by the total frequency ie, $\frac{\sum fx}{\sum f}$ is the first moment about the origine and is denoted by

$$\mu_1 = \frac{\sum fx}{\sum f}$$

$$\mu_2 = \frac{\sum fx^2}{\sum f}$$

$$\mu_3 = \frac{\Sigma f x^3}{\Sigma f}$$

In general, $\mu_r = \frac{\Sigma f \times x^r}{\Sigma f}$ is the r^th moment about origin.

Moment about Mean

The r^{th} moment of x about the mean $\bar{x}$ is denoted by μ_r and is defined as

$$\mu_r = \frac{\Sigma f(x - \bar{x})^r}{\Sigma f}, \qquad r = 0,1,2 \dots . r$$

If $r = 0$, $\mu_0 = 1$, If $r = 1$, $\mu_1 = \frac{\Sigma f(x-\bar{x})}{\Sigma f} = 0$ since the sum of the deviations of the observations about mean is zero

If $r = 2$, $\mu_2 = \frac{\Sigma f(x-\bar{x})^2}{\Sigma f}$,

If $r = 3$, $\mu_3 = \frac{\Sigma f(x-\bar{x})^3}{\Sigma f}$,

If $r = 4$, $\mu_4 = \frac{\Sigma f(x-\bar{x})^4}{\Sigma f}$

ie, $\mu_1 = 0$, $\mu_2 = \frac{\Sigma f d^2}{\Sigma f}$, $\mu_3 = \frac{\Sigma f d^3}{\Sigma f}$, $\mu_4 = \frac{\Sigma f d^4}{\Sigma f}$ where $d = x - \bar{x}$

Moment about Arbitrary point A

For discrete series

The r^th moment of x about the arbitrary point A is denoted by μ_r' and is defined as

$$\mu_r' = \frac{\Sigma f(x - A)^r}{\Sigma f}, \qquad r = 0,1,2 \dots . r$$

$$\mu_0' = 1$$

$$\mu_1' = \frac{\Sigma f(x - A)}{\Sigma f} = \bar{x} - A$$

$$\mu_2' = \frac{\Sigma f(x - A)^2}{\Sigma f}$$

$$\mu_3' = \frac{\Sigma f(x - A)^3}{\Sigma f}$$

$$\mu_4' = \frac{\Sigma f(x - A)^4}{\Sigma f}$$

Relation between Moments about Mean and Moments about arbitrary Point 'A'

$$\mu_1 = 0,$$

$$\mu_2 = \mu_2' - \mu_1'^2$$

$$\mu_3 = \mu_3' - 3\ \mu_2'\ \mu_1' + 2\ \mu_1'^3$$

$$\mu_4 = \mu_4' - 4\ \mu_3'\ \mu_1' + 6\ \mu_1'^2 - 3\ \mu_1'^4$$

For Continuous series

Let $d = \frac{X-A}{C}$ where C is the class intervel

$$\mu_1' = \frac{\sum fd}{\sum f} \times C$$

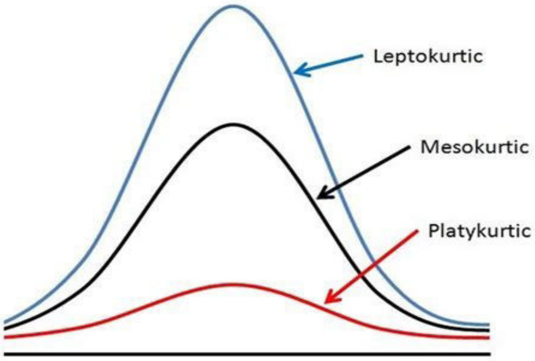
$$\mu_2' = \frac{\sum fd^2}{\sum f} \times C^2$$

$$\mu_3' = \frac{\sum fd^3}{\sum f} \times C^3$$

$$\mu_4' = \frac{\sum fd^4}{\sum f} \times C^4$$

3.3.2 Kurtosis

Besides average, variance and skewness the fourth characteristics used for description and comparison of frequency distribution is the peaked-ness of the distribution. Measure of peaked-ness is called the measure of kurtosis.



Kurtosis refers to the degree of flatness or peaked-ness in the region above the mode of the frequency curve. The degree of kurtosis of a distribution is measured relative to the peaked-ness of the normal curve. For a normal curve the distribution is symmetric about the mean, half the values fall below the mean and half above the mean

The degree of Kurtosis of a distribution is measured relative to the peaked-ness of the normal curve. The measure of Kurtosis tells us the extent to which a distribution is more peaked or flat topped than the normal curve. If the curve is more peaked than the normal curve, it is leptokurtic. On the other hand, if the curve is flatter topped than the normal curve, it is platykurtic. The normal curve is mesokurtic.

Measure of Kurtosis

The most important Measure of Kurtosis is the value of the coefficient β_2. It is defined as

$$\beta_2 = \frac{\mu_4}{{\mu_2}^2}$$

Where μ_4 is the fourth moment about mean and μ_2 is the second moment about mean.

For normal curve $\beta_2 = 3$.

$\beta_1 = \frac{{\mu_3}^2}{{\mu_2}^3}$ is the skewness

Illustration 3.3.4

Find the Kurtosis from the following table

Value :	2	3	4	5	6
Frequency:	1	3	7	3	1

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{60}{15} = 4$$

$$\mu_1 = \frac{\sum fd}{\sum f} = 0 \;, \quad \mu_2 = \frac{\sum fd^2}{\sum f} = \frac{14}{15} = 0.93, \quad \mu_3 = \frac{\sum fd^3}{\sum f} = \frac{0}{15} = 0,$$

$$\mu_4 = \frac{\sum fd^4}{\sum f} = \frac{38}{15} = 2.5$$

$$\beta_2 = \frac{\mu_4}{{\mu_2}^2} = \frac{2.5}{.93} = 2.69$$

Illustration -3.3.5

Find the Kurtosis from the following table

Value :	0-10	10-20	20-30	30-40
Frequency:	1	3	4	2

Solution

	f	Mid x	xf	$d = x - 22$	fd	fd^2	fd^3	fd^4
0-10	1	5	5	-17	-17	289	-4913	83521
10-20	3	15	45	-7	-21	147	-1029	7203
20-30	4	25	100	3	12	36	108	324
30-40	2	35	70	13	26	338	4394	57122
Total	10	60	220		0	810	-1440	148170

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{220}{10} = 22$$

$$\mu_1 = \frac{\sum fd}{\sum f} = 0 \;, \quad \mu_2 = \frac{\sum fd^2}{\sum f} = \frac{810}{10} = 81, \quad \mu_3 = \frac{\sum fd^3}{\sum f} = -\frac{1440}{10} = -144,$$

$$\mu_4 = \frac{\sum fd^4}{\sum f} = \frac{148170}{10} = 148170$$

$$\beta_2 = \frac{\mu_4}{{\mu_2}^2} = \frac{148170}{81^2} = \frac{14817}{6561} = 2.25$$

Illustration 3.3.6

Calculate the first four moments about the mean for the following data and comment on the nature of the distribution :

x	1	2	3	4	5	6	7	8	9
f	1	6	13	25	30	22	9	5	2

Solution

Moment about the point A = 5

x	f	$d = x - 5$	fd	fd^2	fd^3	fd^4
1	1	-4	-4	16	-64	256
2	6	-3	-18	54	-162	486
3	13	-2	-26	52	-104	208

4	25	-1	-25	25	-25	25
5	30	0	0	0	0	0
6	22	1	22	22	22	22
7	9	2	18	36	72	144
8	5	3	15	45	135	405
9	2	4	8	32	128	512
Total	113		-10	282	2	2058

$$\mu_1' = \frac{\sum fd}{\sum f} = -\frac{10}{113} = -0.0885$$

$$\mu_2' = \frac{\sum fd^2}{\sum f} = \frac{282}{113} = 2.4956$$

$$\mu_3' = \frac{\sum fd^3}{\sum f} = \frac{2}{113} = 0.0177$$

$$\mu_4' = \frac{\sum fd^4}{\sum f} = \frac{2058}{113} = 18.2124$$

$$\mu_1 = 0, \quad \mu_2 = 2.4956 - (-0.0885)^2 = 2.4956 - 0.0078 = 2.4878$$

$$\mu_3 = \mu_3' - 3\ \mu_2'\ \mu_1' + 2\ \mu_1'^{\,3} = 0.0177 - 3 \times 2.4956 \times (-0.0885) + 2 \times (-0.0885)^3$$

$$= 0.6789$$

$$\mu_4 = \mu_4' - 4\ \mu_3'\ \mu_1' + 6\ \mu_1'^{\,2} - 3\ \mu_1'^{\,4} = 18.2124 - 4 \times (0.0177) \times (-0.0885) +$$

$$6 \times 2.4956 \times (-0.0885)^2 - 3 \times (-0.0885)^4$$

$$= 18.3357$$

$$\beta_2 = \frac{\mu_4}{{\mu_2}^2} = \frac{18.3357}{(2.4878)^2} = 2.9626$$

Since $\beta_2 < 3$, the given curve is slightly platykurtic.

3.3.3 Lorenz Curve

The Lorenz curve, devised by Max. O. Lorenz, a famous economic statistician, is a graphic method of studying the dispersion in a distribution. This curve was used for the measurement of economic inequalities such as in the distribution of income and wealth between different countries or between different periods of time. Now, the curve is also used in business to study the disparities of the distribution of wages, profits, turnover, production, population, etc. It is a commutative percentage curve in which the percentage of items is combined with the percentage of other things as wealth, profit, turnover etc.

Procedure for Drawing Lorenz Curve

1. The size of the item (variable value) and the frequencies are both cumulated. Taking grand total for each as 100, express these cumulated totals for the variable and the frequencies as percentages of their corresponding grand totals.

2. X-axis representing the percentages of the cumulated frequencies (x) and Y-axis representing the percentages of the cumulated values of the variable (y). Both x and y take the values from 0 to 100

3. Draw the diagonal line y = x joining the origin O (0, 0) with the point P (100, 100) as shown in the diagram. The line OP will make an angle of 45° with the X-axis and is called the line of equal distribution.

4. Plot the percentages of the cumulated values of the variable (y) against the percentages of the corresponding cumulated frequencies (x) for the given distribution and join these points with a smooth freehand curve. If two curves of distribution are shown on the same Lorenz presentation, the curve that is farthest from the diagonal line represents the greeter inequality.

Illustration 3.3.7

In the table below is given the number of companies belonging to two areas A and B according to the amount of profits earned by them. Draw Lorenz curve and interpret them

Profit earned in Rs.'000		6	25	60	84	105	150	170	400
No. of companies	Area A	6	11	13	14	15	17	10	14
	Area B	2	38	52	28	38	26	12	4

Solution

Profits			Area A			Area B		
Profit earned in Rs.'000	cumulative profit	cumulative profit	No. of companies	cumulative companies	cumulative companies	No. of companies	cumulative companies	cumulative companies
6	6	0.6	6	6	6	2	2	1
25	31	3.1	11	17	17	38	40	20

60	91	9.1	13	30	30	52	92	46
84	175	17.5	14	44	44	28	120	60
105	280	28	15	59	59	38	158	79
150	430	43	17	76	76	26	184	92
170	600	60	10	86	86	12	196	98
400	1000	100	14	100	100	4	200	100

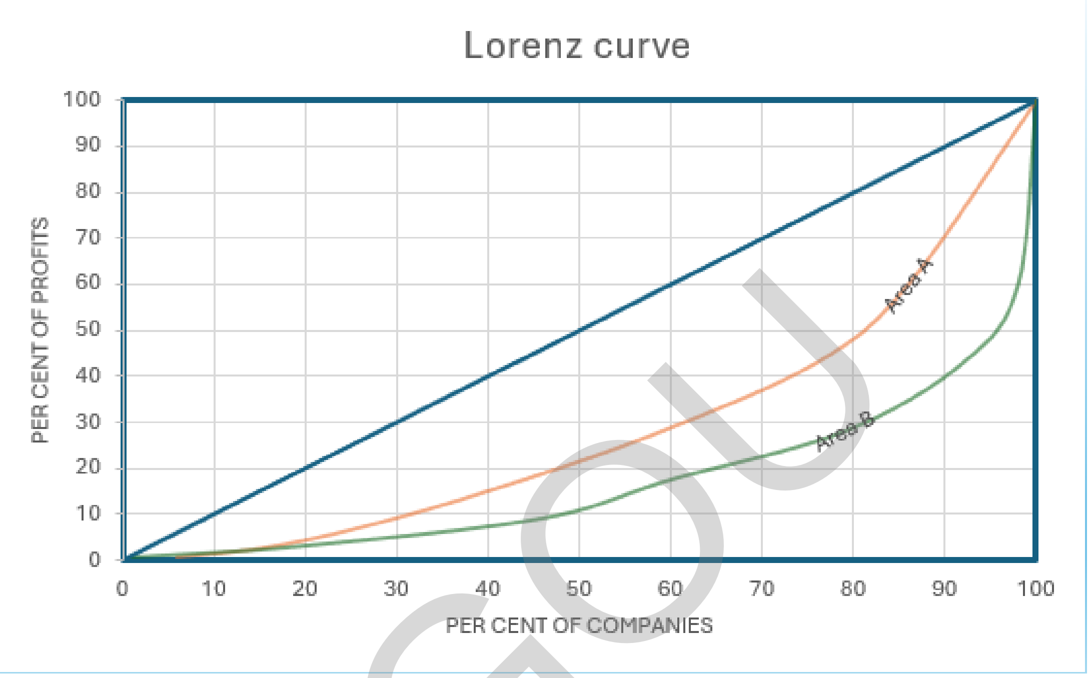


Fig 3.3.1

Since curve B farthest from the diagonal line it represents greater inequality.

3.3.4 Gini Coefficient

The Gini coefficient is a measure of economic inequality within a population. It is commonly used to quantify the distribution of income or wealth among the residents of a country. The coefficient is expressed as a number between 0 and 1, where 0 represents perfect equality (everyone has the same income or wealth) and 1 represents perfect inequality (one person has all the income or wealth, while others have none).

In other words, a higher Gini coefficient indicates a more unequal distribution of income or wealth. The formula for calculating the Gini coefficient involves plotting the Lorenz curve, which represents the cumulative income or wealth shares of the population. The Gini coefficient is then calculated as the area between the Lorenz curve and the line of perfect equality, divided by the total area under the line of perfect equality.

$$\text{Gini coefficient} = \frac{\text{area between the Lorenz curve and the line of perfect equality}}{\text{total area under the line of perfect equality}}$$

Governments, policymakers, and researchers use the Gini coefficient to assess and compare income or wealth inequality across different regions and over time. A lower Gini coefficient is generally associated with a more equal distribution of resources.

Summarised Overview

Skewness and kurtosis are statistical measures that describe the shape of a data distribution. Skewness indicates the degree and direction of asymmetry in a distribution. A distribution with zero skewness is perfectly symmetrical, while a positive or negative value shows the extent to which the distribution leans toward the right or left. Kurtosis, on the other hand, reflects the "tailedness" of a distribution. It measures the concentration of values in the tails versus the center. A normal distribution has a kurtosis of 3 (mesokurtic), while a kurtosis greater than 3 indicates a leptokurtic (peaked) distribution and less than 3 indicates a platykurtic (flatter) one. Kurtosis peaked-ness or flatness in the distribution. Measure of Kurtosis is the value of the coefficient $\beta_2 = \frac{\mu_4}{{\mu_2}^2}$.

The Lorenz curve and Gini coefficient are tools used to measure income or wealth inequality. The Lorenz curve is a graphical representation of the cumulative distribution of income or wealth, plotting the percentage of total income earned against the cumulative percentage of the population. The further the curve lies from the diagonal line of equality, the greater the inequality. The Gini coefficient quantifies this inequality and ranges from 0 (perfect equality) to 1 (maximum inequality). These measures are widely applied in economics and social sciences to assess the fairness of distributions across populations. Lorenz Curve is a commutative percentage curve used for the measurement of economic inequalities. Gini coefficient is a measure of economic inequality within a population.

Assignments

1. Calculate Coefficient of Skewness from the data given below :

x	40-50	50-60	60-70	70-80	80-90	90-100	100-110	110-120	120-130	130-140
f	5	6	8	10	25	30	36	50	60	70

2. Find the variance, skewness and kurtosis of the following distribution

X	0-10	10-20	20-30	30-40
F	1	4	3	2

3. Karl Pearson's coefficient of skewness of a distribution is 0.32. Its standard deviation is 6.5 and mean is 29.6. Find mode and median of the distribution.

4. You are given below the following details relating to the wages is respect of two factories from which it is concluded that the skewness and variability are same in both the factories

	Factory A	Factory B
Mean	50	45
Median	45	50
Standard Deviation	10	10

Point out the mistake or the wrong inference in the above statement.

Calculate the first four moments about the mean for the following data and comment on the nature of the distribution:

X	25	35	45	55	65	75	85
F	5	14	20	25	17	11	8

5. From the following table giving data regarding income of employees in two factories, draw a graph (Lorenz Curve) to show which factory has greater inequalities of income :

Income in Rs.'000	Below 200	200-500	500-1000	1000-2000	2000-3000
Factory A	7000	1000	1200	800	500
Factory B	800	1200	1500	400	200

References

1. Gujarathi , D.&Sangeetha, N. (2007). *Basic Econometrics* (4thed) New Delhi: McGraw Hill

2. Koutsoyianis, A. (1977). *Theory of Econometrics* (2nded). London. The Macmillian Press ltd

Suggested Readings

1. Anderson, D., D.Sweeney and T.Williams (2013): "*Statistics for Business and Economics*", Cengage Learning : New Delhi.

2. Goon, A.M. , Gupta and Das Gupta B (2002): *Fundamentals of Statistics* (Vol I), World Press.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

UNIT 4

Correlation and Regression Analysis

Learning Outcomes

After going through the unit, the learner will be able to:

- understand the concepts of correlation
- compute correlation coefficient between two variables
- understand and develop the statistical applications and analytical skills of data using correlation
- evaluate the uses of correlation and regression analysis in decision making

Background

In real life, we encounter situations that involve the analysis of two or more variables. For example, in mathematics we say that the circumference and the radius of a circle are related by the equation, $C = 2\pi r$ where r and C are variables called independent and dependent variable respectively. In this equation, r can take any value independently, while the value of C varies based on the radius r. In other words, the value of C is depended on the value of r.

That is, in the algebraic formula $Y = aX + b$, X is an independent variable and Y is the dependent variable. For each value of X we can forecast the value of Y using this formula.

Furthermore, there are situations where we may not be in a position to give above type of formulas but can draw comparisons using real life instances. Consider another example that child's age and height. while knowing a child's age may not enable us to precisely predict their height, it does facilitate a more accurate height estimation or forecasting the height with reduced error. This is done through the analysis of the relationship between the variables, age and height, which is termed as Bi-variate analysis. For instance, studying the relationship between income and expenditure of a group of families involve bivariate analysis. The measure of association is called Correlation.

Imagine a scenario where a marketing analyst is meticulously studying the correlation between advertising expenditures and product sales. Your calculated correlation coefficient has indicated a positive relationship between these variables. However, it is important to note that correlation alone does not unveil the intricate mechanisms at play. Regression analysis takes you to the next level by allowing you to construct a model that estimates how alterations in ad spending result in corresponding changes in sales. This model, in turn, transforms into a valuable instrument for refining advertising strategies with the ultimate goal of achieving the highest possible impact on sales outcomes.

This is where regression analysis comes into play, taking you a step beyond correlation. With regression analysis, you can construct a comprehensive model that not only captures the relationship but also provides insight into how changes in ad spending correspond to changes in sales. This model serves as a powerful tool, allowing you to estimate and understand the impact of varying advertising strategies on sales outcomes.

In essence, regression analysis empowers you to optimise advertising strategies for achieving maximum impact on sales. By delving into the underlying dynamics, you gain the ability to make informed decisions that can potentially reshape marketing approaches and enhance overall sales performance.

Keywords

Correlation, Karl-pearson, Rank correlation, Regression

Discussion

3.4.1 Correlation

If two quantities vary in a manner where changes in one are accompanied by changes in the other, then these quantities are said to be correlated. In other words, an increasing in one variable corresponds to an increase or decrease in the other variable, and similarly, a decreasing in one variable corresponds to a decrease or increase in the other variable. When such patterns are observed, the variables are said to be correlated. For instance, income and expenditure are correlated.

3.4.2 Karl Pearson's Coefficient of Correlation

This method is the most widely used method for measuring correlation. It is popularly known as Pearson Ian coefficient of correlation. It is denoted by the symbol "r". It is also known as Product Moment Method.

Computation of correlation coefficient

$$r(x,y) = \frac{Cov\ (x,y)}{\sigma(x)\sigma(y)}$$

Where $Cov\ (x,y)$ = covariance of (x,y). Covariance is a statical measure that quantifies the degree to which two variables change together. It is the sum of the product of the average of the observations from arithmetic mean.

$$Cov\ (x,y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n}$$

$$\sigma(x) = \sqrt{\frac{\Sigma(x-\overline{x})^2}{n}}$$ is the standard deviation of x.

$$\sigma(y) = \sqrt{\frac{\Sigma(y-\overline{y})^2}{n}}$$ is the standard deviation of y.

$$\bar{x} = \frac{\Sigma x}{n}, \qquad \bar{y} = \frac{\Sigma y}{n},$$

So, $$r(x,y) = \frac{\frac{\Sigma(x-\bar{x})(y-\bar{y})}{n}}{\sqrt{\frac{\Sigma(x-\overline{x})^2}{n}}\sqrt{\frac{\Sigma(y-\overline{y})^2}{n}}} = \frac{\Sigma(x-\bar{x})\ (y-\bar{y})}{\sqrt{\Sigma(x-\overline{x})^2}\ \sqrt{\Sigma(y-\overline{y})^2}}$$

The above formula can be expressed in the following form also

$$r(x,y) = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

Illustration 3.4.1

Find the Karl Pearson's coefficient between x and y for the following data

$$n = 10, \quad \sum x = 35, \quad \sum x^2 = 203, \sum y = 28, \sum y^2 = 140, \sum xy = 168$$

Solution

$$r(x,y) = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

$$= \frac{10 \times 168 - 35 \times 28}{\sqrt{10 \times 203 - 35^2}\sqrt{10 \times 140 - 28^2}}$$

= 0.99

Illustration 3.4.2

Find Karl Pearson's correlation coefficient between x and y for the following data,

$$n = 15, \quad Cov(x,y) = 8.13, \sigma(x) = 3.01, \quad \sigma(y) = 3.03$$

Solution

$$r = \frac{Cov(x,y)}{\sigma(x)\sigma(y)} = \frac{8.13}{3.01 \, x \, 3.03} = 0.89$$

Illustration 3.4.3

Find Karl Pearson's correlation coefficient between x and y for the following data,

$$n = 1000, \ \sigma(x) = 4.5, \quad \sigma(y) = 3.6, \quad \sum(x - \bar{x}).(y - \bar{y}) = 4800$$

Solution

$$r(x,y) = \frac{Cov\ (x,y)}{\sigma(x)\sigma(y)}$$

$$r = \frac{\frac{\Sigma(x-\bar{x})(y-\bar{y})}{n}}{\sigma(x)\sigma(y)}$$

$$= \frac{\frac{4800}{1000}}{4.5 \times 3.6}$$

$$= \frac{4.8}{16.2}$$

$$= 0.296$$

Illustration 3.4.4

Find Product Moment method of Correlation Coefficient between x and y for the following data,

$$n = 20, \Sigma(x_i - \bar{x})^2 = 136, \Sigma(y_i - \bar{y})^2 = 138, \ \sum(x - \bar{x}).(y - \bar{y}) = 122$$

Solution

$$r = \frac{\Sigma(x-\bar{x})\ (y-\bar{y})}{\sqrt{\Sigma(x-\overline{x})^2}\ \sqrt{\Sigma(y-\overline{y})^2}}$$

$$= \frac{122}{\sqrt{136}\sqrt{138}}$$

$$= \frac{122}{11.66 \times 11.75}$$

$= 0.89$

Illustration 3.4.5

Calculate the coefficient of correlation for the following table by Karl Pearson's coefficient of correlation.

X	1	2	3	4	5
Y	3	1	2	5	4

Solution

$$\bar{x} = \frac{\sum x}{n} = \frac{15}{5} = 3$$

$$\bar{y} = \frac{\sum y}{n} = \frac{15}{5} = 3$$

x	y	$x-5$	$y-5$	$(x-5)^2$	$(y-5)^2$	$(x-5)(y-5)$
1	3	-2	0	4	0	0
2	1	-1	-2	1	4	2
3	2	0	-1	0	1	0
4	5	1	2	1	4	2
5	4	2	1	4	1	2
				10	10	6

$$r = \frac{\Sigma(x-\bar{x})\,(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\,\sqrt{\Sigma(y-\bar{y})^2}} = \frac{6}{\sqrt{10}\,\sqrt{10}} = \frac{6}{10} = 0.6$$

Illustration 3.4.6

Calculate the coefficient of correlation for the following table by Karl Pearson's coefficient of correlation.

X	70	69	68	67	66	65	65
Y	72	68	70	68	65	66	66

Solution

$$\bar{x} = \frac{\sum x}{n} = \frac{469}{7} = 67$$

$$\bar{y} = \frac{\sum y}{n} = \frac{476}{7} = 68$$

x	y	$x-67$	$y-68$	$(x-67)^2$	$(y-68)^2$	$(x-67)(y-68)$
70	72	3	4	9	16	12
69	68	2	0	4	0	0
68	70	1	2	1	4	2
67	68	0	0	0	0	0
66	65	-1	-3	1	9	3
65	67	-2	-1	4	1	2
64	66	-3	-2	9	4	6
469	476			28	34	25

$$r = \frac{\Sigma(x-\bar{x})\ (y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\ \sqrt{\Sigma(y-\bar{y})^2}}$$

$$= \frac{25}{\sqrt{28}\ \sqrt{34}}$$

$$= \frac{25}{30.85}$$

$$= 0.81$$

Illustration 3.4.7

Calculate the coefficient of correlation for the following table by Karl Pearson's coefficient of correlation.

X	9	8	10	12	11	13	14	16	15
Y	1	2	3	4	5	6	7	8	9

Solution

$$\bar{x} = \frac{\sum x}{n} = \frac{108}{9} = 12$$

$$\bar{y} = \frac{\sum y}{n} = \frac{45}{9} = 5$$

x	y	$x-67$	$y-68$	$(x-67)^2$	$(y-68)^2$	$(x-67)(y-68)$
9	1	-3	-4	9	16	12
8	2	-4	-3	16	9	12

10	3	-2	-2	4	4	4
12	4	0	-1	0	1	0
11	5	-1	0	1	0	0
13	6	1	1	1	1	1
14	7	2	2	4	4	4
16	8	4	3	16	9	12
15	9	3	4	9	16	12
108	45			60	60	57

$$r = \frac{\Sigma(x-\bar{x})\,(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\,\sqrt{\Sigma(y-\bar{y})^2}}$$

$$= \frac{57}{\sqrt{60}\,\sqrt{60}}$$

$$= \frac{57}{60}$$

$$= 0.95$$

Illustration 3.4.8

Find Karl Pearson's coefficient of correlation, from the following series of marks secured by ten students in a class test in mathematics and statistics.

Marks in Mathematics	45	70	65	30	90	40	50	75	85	60
Marks in Statistics	35	90	70	40	95	40	60	80	80	50

a. Also calculate its probable error.

b. Hence discuss if the value of "r" is significant or not? Also compute the limits within which the population correlation coefficient may be expected to lie?

Solution

x	y	$x-61$	$y-64$	$(x-61)^2$	$(y-64)^2$	$(x-61)(y-64)$
45	35	-16	-29	256	841	464
70	90	9	26	81	676	234
65	70	4	6	16	36	24
30	40	-31	-24	961	576	744
90	95	29	31	841	961	899

40	40	-21	-24	441	576	504
50	60	-11	-4	121	16	44
75	80	14	16	196	256	224
85	80	24	16	576	256	384
60	50	-1	-14	1	196	14
Total 610	640			3490	4390	3535

$$\bar{x} = \frac{\sum x}{n} = \frac{610}{10} = 61$$

$$\bar{y} = \frac{\sum y}{n} = \frac{640}{10} = 64$$

$$r = \frac{\Sigma(x-\bar{x})\,(y-\bar{y})}{\sqrt{\Sigma(x-\overline{x})^2}\,\sqrt{\Sigma(y-\overline{y})^2}} = \frac{3535}{\sqrt{3490}\,\sqrt{4390}} = \frac{3535}{59.08 * 66.26} = 0.903.$$

Illustration 3.4.10

The following data relates to the percentage of failures in the Higher Secondary examination. Find out whether there is any correlation between age and failure in the examination.

Age of Candidates (in Years)	13	14	15	16	17	18	19	20	21
Percentage of Failure	39	40	43	34	36	39	48	47	52

Solution

x	y	$x-17$	$y-42$	$(x-17)^2$	$(y-42)^2$	$(x-17)(y-42)$
13	39	-4	-3	16	9	12
14	40	-3	-2	9	4	6
15	43	-2	1	4	1	-2
16	34	-1	-8	1	64	8
17	36	0	-6	0	36	0
18	39	1	-3	1	9	-3
19	48	2	6	4	36	12
20	47	3	5	9	25	15
21	52	4	10	16	100	40
Total-153	378			60	284	88

$$\bar{x} = \frac{\sum x}{n} = \frac{153}{9} = 17$$

$$\bar{y} = \frac{\sum y}{n} = \frac{378}{9} = 42$$

$$r = \frac{\sum(x-\bar{x})\,(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\,\sqrt{\sum(y-\bar{y})^2}} = \frac{88}{\sqrt{60}\,\sqrt{284}}$$

$$= \frac{88}{7.75 \times 16.85}$$

$$= \frac{88}{130.6}$$

$= 0.67$

3.4.3 Rank Correlation

Karl Pearson's Coefficient of correlation is purely based on magnitudes of the variables. However, there are situations to measure qualitative data rather than quantitative one. For instance, to measure qualitative data such as intelligence, honesty, character etc., it is feasible to find ranks for individuals. The correlation coefficient derived from these ranks is called Rank correlation coefficient.

In other words, the Rank Correlation Coefficient between two variables is a correlation coefficient obtained based on the ranking of the variables. Edward Spearman in 1904 devised a formula known as Spearman's Rank correlation Coefficient. Spearman's Rank correlation Coefficient.

$$r = 1 - 6\frac{\sum D^2}{n(n^2-1)}$$

Where, r is the Rank correlation coefficient.

D is the difference of the corresponding ranks.

n is the number of items.

In rank correlation coefficient we have to do two types of problems

1. When actual ranks are given,
2. When ranks are not given.

When Actual Ranks are Given

♦ If the actual ranks are given, the steps required for computing Spearman's Correlation Coefficient are;

- Take the differences of the two ranks, that is (R_{1-} R_2) and denote these differences by D
- Square these differences and obtain the total $\sum$ D^2
- Apply the formula $r = 1 - 6\frac{\sum D^2}{n(n^2-1)}$

Illustration.3.4.11

Following are the marks obtained by 10 students of two subjects Mathematics and Physics in a class test. Estimate Spearman's Rank Correlation.

Name of students	A	B	C	D	E	F	G	H
Mathematics	7	3	1	4	6	8	2	5
Physics	6	2	1	5	8	7	3	4

Solution

Name of students	Mathematics	Physics	D	D^2
A	7	6	1	1
B	3	2	1	1
C	1	1	0	0
D	4	5	1	1
E	6	8	2	4
F	8	7	1	1
G	2	3	1	1
H	5	4	1	1
				10

$$r = 1 - 6\frac{\sum D^2}{n(n^2-1)}$$

$$= 1 - 6\frac{10}{8(8^2-1)}$$

$$= 1 - \frac{60}{504}$$

$$= 1 - 0.12$$

$$= 0.88$$

When Ranks are Not Given

When we are given the actual data rather than the ranks, it become necessary to assign ranks. Ranks can be assigned by taking either the highest value as the first rank or the lowest value as the first rank. However, the same method must be consistently applied to both variables.

Illustration 3.4.12

The data given below relates to the price and demand of a commodity over a period. Compute the correlation coefficient between the Price and Demand

Price	50	75	60	70	95	90	88
Demand	100	140	110	115	150	134	120

Solution

Price	Demand	Rank of Price R_1	Rank of Demand R_2	Difference D	D^2
50	100	7	7	0	0
75	140	4	2	2	4
60	110	6	6	0	0
70	115	5	5	0	0
95	150	1	1	0	0
90	134	2	3	1	1
88	120	3	4	1	1
Total					6

$$r = 1 - 6\frac{\sum D^2}{n(n^2-1)}$$

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

$$= 1 - \frac{36}{336}$$

$$= 1 - 0.107$$

$$= 0.893$$

Illustration 3.4.13

Find the Spearman's rank correlation coefficient between marks in accountancy and statistics

Marks in Statistics	48	60	72	62	56	40	39	52	30
Marks in Accountancy	62	78	65	70	38	54	60	32	31

Solution

Marks in Statistics	Marks in Accountancy	R_1	R_2	D	D^2
48	62	6	4	2	4
60	78	3	1	2	4
72	65	1	3	2	4
62	70	2	2	0	0
56	38	4	7	3	9
40	54	7	6	1	1
39	60	8	5	3	9
52	32	5	8	3	9
30	31	9	9	0	0
Total					40

$$r = 1 - 6\frac{\sum D^2}{n(n^2-1)}$$

$$= 1 - 6\ \frac{40}{9(9^2-1)}$$

$$= 1 - \frac{240}{720}$$

$$= 0.6667$$

In Case the Ranks are Equal

In certain instances, we might encounter cases where two or more items share equal ranks. In such cases, each individual item is assigned an average rank. For example, if two values are both ranked equal at the third place, each is given a rank of = $\frac{3+2}{2} = 3.5$

However, when three values are ranked equally at the third place, the individual ranks are calculated as = $\frac{3+4+5}{3} = 4$. When equivalent ranks are assigned to multiple entries, certain adjustments in the formula become necessary for calculating the Rank Correlation coefficient. This adjustment involves adding $\frac{m^3-m}{12}$ to the sum of squared differences $\sum D^2$,where "m" stands for the number of items which have the common rank. In case, there are more than one such group of items with same rank, the value is added as many times as the number of such groups. The formula in that case is written as

$$r \;= 1 - \frac{6[(\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \ldots\ldots]}{n(n^2 - 1)} + \cdots$$

Illustration 3.4.14

Calculate the coefficient of rank correlation from the following data

X	48	33	40	9	16	16	65	24	16	57
Y	13	31	31	6	15	4	20	9	6	19

Solution

Ranks are assigned as follows for x series;

65 = The highest value, so first rank, 57 = Gets the second rank, 48 = Gets the third rank

40 = Gets the fourth rank, 33 = Gets the fifth rank, 24 = Gets the sixth rank

Now the next highest value 16 is repeated thrice, therefore average of the next three ranks will be taken, that is $\frac{7+8+9}{3} = 8^{th}\ th$ rank. So, rank 8 will be assigned to all the values of 16. The last value 9 gets the 10th rank.

Now, let us explain how the ranks for y series are assigned.

The highest value 31 is repeated twice, so the respective ranks will be $\frac{1+2}{2}$ = 1.5

Next value 6 is repeated twice, so the next two ranks will be averaged and assigned, that is, $\frac{8+9}{2}$ = 8.5.

x	R_1	Y	R_2	$D=R_{1-}R_2$	D^2
48	3	13	6	3	9
33	5	31	1.5	3.5	12.25
40	4	31	1.5	2.5	6.25
09	10	6	8.5	1.5	2.25
16	8	15	5	3	9
16	8	4	10	2	4
65	1	20	3	2	4
24	6	9	7	1	1
16	8	6	8.5	.05	0.25
57	2	19	4	2	4
Total					**52**

Rank Correlation Coefficient is calculated using the equation;

$$r \quad = 1 - \frac{6[(\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \dots\dots]}{n\,(n^2 - 1)} + \cdots$$

$$= 1 - \frac{6[(52 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)]}{10^3 - 10}$$

$$= 1 - \frac{6(52+2+0.5+0.5)}{990}$$

$$= 1 - \frac{6 \times 55}{990}$$

= 0.667

Illustration 3.4.15

Eight students have obtained the following marks in Economics and Accountancy. Calculate the rank correlation coefficient

Marks in Accountancy	25	30	38	22	50	70	30	90
Marks in Economics	50	40	60	40	30	20	40	70

Solution

Computation is explained in the following table.

X	R_1	Y	R_2	$D = R_1 - R_2$	D^2
25	7	50	3	4	16
30	5.5	40	5	0.05	0.25
38	4	60	2	2	4
22	8	40	5	3	9
50	3	30	7	4	16
70	2	20	8	6	36
30	5.5	40	5	0.05	0.25
90	1	70	1	0	0
Total					**81.5**

Here two correction factors are to be added to the equation, for X series, 30 is repeated twice, so the correction factor $\frac{2^3-2}{12}$ is added. Similarly for Y

Series value 40 is repeated thrice, so the correction factor $\frac{3^3-3}{12}$is added.

The rank correlation coefficient can be calculated using the equation;

$$r \quad = 1 - \frac{6[(\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \ldots\ldots]}{n\,(n^2 - 1)} + \cdots$$

$$= 1 - \frac{6[(81.5+\frac{1}{12}(2^3-2)+\frac{1}{12}(3^3-3)]}{8^3-8}$$

$$= 1 - \frac{6[(81.5+0.5+2]}{8^3-8}$$

$$= 1 - \frac{6 \times 84}{504}$$

$$= 1 - 1 = 0$$

3.4.4 Regression

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

In regression analysis there are two types of variables. The variable whose value is influenced or is to be predicted is called dependent variable and the variable which influences the values or is used for prediction is called independent variable. In regression analysis, independent variable is also known as regressor or predictor or explanatory variable while the dependent variable is also known as regressed or explained variable.

3.4.4.1 Methods for Studying Regression

The following methods are used for studying regression

Freehand Method:

This is not a scientific method. In this method, a graph is plotted based on the variables, by taking time on X axis and the other variable on Y axis. Then we join the plots on the graph using a freehand trend line. Thus, the name freehand method. If the plotted lines can form a straight line, we interpret it as linear regression and if it is a curve, we interpret it as a nonlinear regression. The method is simple, quick and easy, but the interpretation is based on the personal assumptions, so based on the limitations of the method, this method is not popular.

Method of Least Squares:

The least-squares regression method is a technique commonly used in Regression Analysis. It is a mathematical method used to find the best fit line that represents the relationship between an independent and dependent variable. It aims to minimise the sum of the squared differences between the observed data and the corresponding values predicted by the model. To understand the least-squares regression method. Let us get familiar with the concepts involved in formulating the line of best fit.

The mathematical expression capturing the connection between an independent variable and a dependent variable takes the form of a linear regression line, often represented as a straight-line equation, $y = ax + b$, where ‘a’ and 'b' are constants. When we are presented with an observed pair (x, y), they signify the values of the related independent and dependent variables.

For each data point, calculate the difference between the observed value and the predicted value derived from the model. This difference is referred to as residuals, essentially capturing the disparity between actual and estimated outcomes.

By applying mathematical optimisation techniques to minimise the sum of squares of these residuals, we reach a simplified yet powerful outcome. This process allows us to uncover the most suitable linear regression line that best fits the data, making the relationship between variables clearer and more predictive.

Apply mathematical optimisation to minimise the sum of squares of the residual, we get the normal equations,

$$\sum x = Na + b\sum y$$

$$\sum xy = a\sum y + b\sum y^2$$

Solving these equations and substituting the values of a and b in the equation $y = ax + b$ we get the equation of the regression lines.

Lines of Regression

If the variables in a bivariate distribution are related, we will find that the points in the scatter diagram will cluster round some curve called the “curve of regression”. In a bivariate distribution where variables are interconnected, the scatter diagram’s points tend to cluster around a curve known as the “curve of regression”. Should this curve adopt a linear shape, it’s termed the “line of regression”, signifying a linear regression between the variables. Conversely, if the curve deviates from a straight line, the regression is termed “curvilinear.”

For the scenario of two variables, x and y, we end up with two regression lines: one for x on y and the other for y on x. The regression line of y on x yields the most probable y values for given x values, while the regression line of x on y provides the most probable x values for given y values. Consequently, we have two regression equations that aid in understanding the predictive relationships between these variables.

Regression Equations

Regression equations are algebraic expression of the regression lines. Since there are two regression lines, there are two regression equations.

The regression equation of x on y is used to describe the variation in the values of

x for given changes in y and the regression equation of y on x is used to describe the variation in the values of y for given changes in x.

Regression Equation y on x

The regression equation y on x is expressed as follows;

$$y = a + b\,x$$

In this equation, "a" and "b" are unknown constants. These constants are called the parameters of the line. The values of "a" and "b" can be obtained by solving the following equations simultaneously.

$$\sum y = Na + b\sum x$$

$$\sum xy = a\sum x + b\sum x^2$$

These equations are commonly referred to as the normal equations. By solving these normal equations and substituting the determined values of 'a' and 'b' into the equation, we obtain the regression equation for predicting 'y' based on 'x'.

Regression Equation of x on y

The regression equation x on y is expressed as follows;

$$x = a + b\,y$$

To determine the values of "a" and "b", the following two normal equations are to be solved simultaneously

$$\sum x = Na + b\sum y$$

$$\sum xy = a\sum y + b\sum y^2$$

Properties of Regression Lines

The regression line is constructed to minimise the sum of squared residuals (the differences between observed and predicted values), ensuring the line provides the best fit to the data.

- The regression lines pass through the mean $(\bar{x}, \bar{y})$ of X and Y variables.
- The two regression lines are perpendicular when $r = 0$.
- The regression line of y on x has the same slope as the regression line of x on y, ensuring symmetry in the relationships between variables.
- The independent variable is not influenced by the errors in the dependent variable, ensuring that causality is properly addressed.
- The regression line can be used to predict values of the dependent variable

based on the known values of the independent variable(s).

- The regression line assumes a linear relationship between the variables. If the true relationship is nonlinear, the regression line might not accurately represent the relationship.

Illustration 3.4.16

From the following data, obtain the two regression equations by the method of least square

x	10	6	10	6	8
y	6	2	10	4	8

Solution

x	y	$x \times y$	x^2	y^2
10	6	60	100	36
6	2	12	36	04
10	10	100	100	100
6	4	24	36	16
8	8	64	64	64
Total-40	30	260	336	220

Regression equation y on x is given by $y = a + b\,x$

To determine the value of constants "a" and "b", the following two normal equations are to be solved;

$$\sum y = Na + b\sum x$$

$$\sum xy = a\sum x + b\sum x^2$$

Substituting the values in the equation, we get;

$$30 = 5a + 40b$$

$$260 = 40a + 336b$$

Multiplying equation 1 by 8 we get;

$$240 = 40a + 320b$$

$260 = 40a + 336b$

Subtracting equation (4) from (3), we get;

$b = \frac{20}{16} = 1.25$

This value of "b" can be substituted in equation (1), we get the value of "a". That is;

$30 = 5a + 40x1.25$

$30 = 5a + 50$

$a = {}^{-}4$

Substituting the values of "a" and "b" in the regression equation, we get the regression line of y on x, $y = 1.25x - 4$

Now we can calculate the regression equation of x on y, that is given by the equation;

$x = a + b\,y$, and the two normal equations are

$\sum x = N\,a + b\sum y$

$\sum xy = a\sum y + b\sum y^2$

Substituting the values in the equation, we get;

$40 = 5a + 30b$

$260 = 30a + 220b$

Multiplying equation (1) by 6, we get

$240 = 30a + 180b$

$260 = 30a + 220b$

$b = \frac{20}{40} = 0.5$

Substituting the value of "b" in equation (1), we get;

$a = \frac{25}{5} = 5$ 5

The regression line of x on y is $x = 5 + 0.5y$

Illustration 3.4.17

From the following data, obtain the two regression equations by the method of least square.

x	6	2	10	4	8
y	9	11	5	8	7

Solution

x	y	$x \times y$	x^2	y^2
6	9	36	81	54
2	11	4	121	22
10	5	100	25	50
4	8	16	64	32
8	7	64	49	56
Total=30	40	220	340	214

Regression equation y on x is given by $y = a + b\,x$

Two normal equations are

$$\sum y = Na + b\sum x$$

$$\sum xy = a\sum x + b\sum x^2$$

Substituting the values in the equation, we get;

$$40 = 5a + 30b$$

$$214 = 30a + 220b$$

Solving the equations we get,

$$a = 11.9, \quad b = -0.65$$

Regression equation y on x is given by $y = 11.9 - 0.65\,x$

Regression equation x on y is given by $x = a + b\,y$

Two normal equations are

$$\sum x = Na + b\sum y$$

$$\sum xy = a\sum y + b\sum y^2$$

Substituting the values in the equation, we get;

$$30 = 50a + 40b$$

$$214 = 40a + 340b$$

Solving the equations we get,

$a = \frac{41}{385}, \quad b = \frac{95}{154}$

Regression equation y on x is given by $y = \frac{41}{385} + \frac{95}{154}\ x$

Deviations Taken from Arithmetic Means of x and y Series

The calculations can be simplified if we take deviations from actual means of x and y series, instead of actual values of x and y.

In such a case, the equation y on x is written as

$$(y - \bar{y}\ ) = b_{yx}\ (x - \bar{x})$$

Where $\bar{y} = \frac{\sum y}{n}, \quad \bar{x} = \frac{\sum x}{n}, \quad b_{yx} = r\frac{\sigma_y}{\sigma_x}$

Similarly, regression equation x on y can be obtained as follows;

$$(x - \bar{x}\ ) = b_{xy}\ (y - \bar{y})$$

Where $\bar{y} = \frac{\sum y}{n}, \quad \bar{x} = \frac{\sum x}{n}, \quad b_{xy} = r\frac{\sigma_x}{\sigma_y}$

Regression Coefficients

Since there are two regression equations, there are two regression coefficients.

The two regression coefficients are regression coefficient of x on y and regression coefficient of y on x.

Regression Coefficient of x on y

The regression coefficient of x on y is represented by the symbol, bxy. This regression coefficient measures the change in x, corresponding to a unit change in y. The regression coefficient of x on y is given by

$$b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

Where, r = Karl Pearson's Correlation Coefficient

σx = Standard deviation of x series

σy = Standard deviation of y series

Regression Coefficient of y on x

The regression coefficient of y on x is represented by byx This regression coefficient measures the change in y variable corresponding to unit change in x variable. The value of byx is given by;

$$b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

where r = Karl Pearson's correlation Coefficient

σy = Standard deviation of y series

σx = Standard deviation of x series

The regression coefficients b_{yx} and b_{xy} can be easily obtained by using the formula

$$b_{yx} = \frac{\Sigma(x-\bar{x})\,(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

$$b_{xy} = \frac{\Sigma(x-\bar{x})\,(y-\bar{y})}{\Sigma(y-\bar{y})^2}$$

Calculating Correlation Coefficients from Regression Coefficients

We know that $b_{xy} = r\frac{\sigma_x}{\sigma_y}$ and $b_{yx} = r\frac{\sigma_y}{\sigma_x}$

Therefore $b_{xy} \times b_{yx} = r\frac{\sigma_x}{\sigma_y} \times r\frac{\sigma_y}{\sigma_x}$

Cancelling the common items, we get $b_{xy} \times b_{yx} = r^2$

Thus, correlation coefficient can be calculated using the equation;

$$r = \sqrt{bxy\ \ x\ \ byx}$$

Since the value of the correlation coefficient cannot exceed one, one of the regression coefficients must be less than one. In other words, both the regression coefficients cannot be greater than one. Similarly, both the regression coefficients will have the same sign, that is they will be either positive or negative.

3.4.5 Properties of Regression Coefficients

- Correlation coefficient is the geometric mean between the regression coefficients
- If one of the regression coefficients is greater than unity. the other must be less than unity.
- Arithmetic mean of the regression coefficients is greater than the correlation coefficient r provided. $r > 0.$

- Regression coefficients are independent of the change of origin but not of scale.
- Both the regression coefficients will have the same sign.
- The sign (positive or negative) of the regression coefficient indicates the direction of the relationship between variables. A positive regression coefficient suggests a positive correlation, while a negative coefficient implies a negative correlation.
- The magnitude of the regression coefficient reflects the strength of the relationship. A larger absolute value indicates a stronger influence of the independent variable on the dependent variable.

Illustration 3.4.18

From the_following_data, obtain the regression equation of x and y and y on x

x	10	6	10	6	8
y	6	2	10	4	8

Solution

x	y	$x-8$	$(x-8)^2$	$y-6$	$(y-6)^2$	$(x-8)(y-6)$
10	6	2	4	0	0	0
6	2	$^-2$	4	$^-4$	16	8
10	10	2	4	4	16	8
6	4	$^-2$	4	$^-2$	4	4
8	8	0	0	2	4	0
Total 40	30	0	16	0	40	20

$n=5,\ \bar{x}=\frac{\Sigma x}{n}=\frac{40}{5}=8,\quad \bar{y}=\frac{\Sigma y}{n}=\frac{30}{5}=6$

Regression equation y on x is given by the equation

$$(y-\bar{y}\ )\ =\ b_{yx}\ (x-\bar{x})$$

$$b_{yx}=\frac{\Sigma(x-\bar{x})\ (y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

$$=\frac{20}{16}=1.25$$

$$(y-6\ )\ =\ 1.25\ (x-8)$$

$$y=1.25x-4$$

Similarly, regression equation x on y is given by the formula

$$(x - \bar{x}\ ) = b_{xy}\ (y - \bar{y})$$

$$b_{xy} = \frac{\Sigma(x - \bar{x})\ (y - \bar{y})}{\Sigma(y - \bar{y})^2}$$

$$= \frac{20}{40} = 0.5$$

$$(x - 8\ ) = 0.5\ (y - 6)$$

$$x = 0.5\ y + 5$$

Illustration 3.4.20

From the following data, obtain the regression equation of x and y and y on x

X	20	22	25	26	27	33
Y	31	29	32	37	35	34

Solution

x	y	$x - 25.5$	$y - 33$	$(x - 25.5)(y - 33)$	$(x - 25.5)^2$	$(y - 33)^2$
20	31	-5.5	-2	11	30.25	4
22	29	-3.5	-4	14	12.25	16
25	32	-0.5	-1	0.5	0.25	1
26	37	0.5	4	2	0.25	16
27	35	1.5	2	3	2.25	4
33	34	7.5	1	7.5	56.25	1
153	198			38	101.5	42

$$n = 6,\ \ \bar{x} = \frac{\Sigma x}{n} = \frac{153}{6} = 25.5,\ \ \bar{y} = \frac{\Sigma y}{n} = \frac{198}{6} = 33$$

Regression equation y on x is given by the equation

$$(y - \bar{y}\ ) = b_{yx}\ (x - \bar{x})$$

$$b_{yx} = \frac{\Sigma(x - \bar{x})\ (y - \bar{y})}{\Sigma(x - \bar{x})^2}$$

$$= \frac{38}{101.5} = 0.37$$

$$(y - 33\ ) = 0.37\ (x - 25.5)$$

$y = 0.37x - 23.57$

Similarly, regression equation x on y is given by the formula

$$(x - \bar{x}\ ) = b_{xy}\ (y - \bar{y})$$

$$b_{xy} = \frac{\Sigma(x - \bar{x})\ (y - \bar{y})}{\Sigma(y - \bar{y})^2}$$

$$= \frac{38}{42} = 0.9$$

$$(x - 25.5\ ) = 0.9\ (y - 33)$$

$$x = 0.9\ y - 4.2$$

Illustration 3.4.21

Find the following from the data

X	25	28	35	32	31	36	29	38	34	32
Y	43	46	49	41	36	32	31	30	33	39

1. The two regression equations
2. The coefficient of correlation between the marks in Economics and Statistics
3. The most likely marks in Statistics when marks in Economics are 30.

Solution

X	y	x-32	y-38	(x-32) (y-38)	$(x-32)^2$	$(y-38)^2$
25	43	-7	5	-35	49	25
28	46	-4	8	-32	16	64
35	49	3	11	33	9	121
32	41	0	3	0	0	9
31	36	-1	-2	2	1	4
36	32	4	-6	-24	16	36
29	31	-3	-7	21	9	49
38	30	6	-8	-48	36	64

34	33	2	-5	-10	4	25
32	39	0	1	0	0	1
320	380			-93	140	398

$n=10$, $\bar{x} = \frac{\sum x}{n} = \frac{320}{10} = 32$, $\bar{y} = \frac{\sum y}{n} = \frac{380}{10} = 38$

Regression equation y on x is given by the equation

$$(y - \bar{y}\ ) = b_{yx}\,(x - \bar{x})$$

$$b_{yx} = \frac{\sum(x-\bar{x})\,(y-\bar{y})}{\sum(x-\bar{x})^2}$$

$$= \frac{-93}{140} = -0.66$$

$$(y - 38\ ) = -0.66\,(x - 32)$$

$$y = -0.66x + 59.12$$

Similarly, regression equation x on y is given by the formula

$$(x - \bar{x}\ ) = b_{xy}\,(y - \bar{y})$$

$$b_{xy} = \frac{\sum(x-\bar{x})\,(y-\bar{y})}{\sum(y-\bar{y})^2}$$

$$= \frac{-93}{398} = -0.2337$$

$$(x - 32\ ) = -0.23\,(y - 38)$$

$$x = -0.23\ y + 40.74$$

2) Coff. of correlation r $= \sqrt{bxy\ \ x\ \ byx}$

$= \sqrt{-0.66\ \ x\ \ -0.23}$

$= \sqrt{0.1518}$

$= \pm 0.389$

3) Regression line of y on x is

$$y = -0.66x + 59.12$$

When x = 30, we get

$y = -0.66 \times 30 + 59.12$

$= 39.32$

Illustration 3.4.22

If the regression equations between the variables x and y are $4x - 5y + 33 = 0$ and $20x - 9y = 107$, find the correlation coefficient and means of the variables.

Solution

The regression equations $4x - 5y + 33 = 0$

$20x - 9y = 107$

Solving we get $x = 13, \ \ y = 17$

Since the regression lines pass through $\overline{(x, y)}$ we have $\bar{x} = 13, \ \bar{y} = 17$

Rewriting the regression lines of y on x $4x - 5y + 33 = 0$ as

$y = \frac{4}{5}x + \frac{33}{5}$, we get $b_{yx} = r\frac{\sigma y}{\sigma x} = \frac{5}{4}$

Similarly, Rewriting the regression lines of x on y, $20x - 9y = 107$ as

$y = \frac{9}{20}y + \frac{107}{9}$, we get $b_{xy} = r\frac{\sigma x}{\sigma y} = \frac{9}{20}$

Thus $r = \sqrt{b_{yx}.b_{xy}} = \sqrt{\frac{5}{4}.\frac{9}{20}} = \pm 0.6$

Since b_{yx} and b_{xy} are positive, $r = 0.6$

Illustration 3.4.23

From the following data, find the most likely value of y when x =24. Given $r = 0.58$

	y	X
Mean	985.8	18.1
S.D	36.4	2.0

Solution

The regression equation of y on x is

$$(y - \bar{y}\ ) = b_{yx}\ (x - \bar{x})$$

$$(y - \bar{y}\ ) = r\frac{\sigma y}{\sigma x}\ (x - \bar{x})$$

Given $\bar{x} = 18.1$, $\bar{y} = 985.8$, $\sigma x = 2$, $\sigma y = 36.4$ and $r = 0.58$.

Substituting these in the equation,

$$(y - 985.8\ ) = \frac{0.58 \times 36.4}{2.0}(x - 18.1)$$

$$(y - 985.8\ ) = 10.556\,(x - 18.1)$$

$$y = 10.556\,x + 794.74$$

When $x = 24$, $y = 10.556 \times 24 + 794.74 = 1048.084$

Summarised Overview

Coefficient of correlation is the degree of relationship between two variables is called coefficient of correlation. Karl Pearson's Coefficient of Correlation refers to a method most widely used method for measuring correlation. Degree of Correlation is used to interpret the computed value of Karl Pearson's Correlation coefficient. The correlation coefficient obtained from the ranks is called Rank correlation coefficient.

The least-squares regression method is a technique commonly used in Regression analysis. Regression equations are algebraic expression of the regression lines. Lines of Regression points in the scatter diagram will cluster round a straight line called line of regression. Regression line of x on y and Regression line of y on x. Regression Coefficients - There are two regression coefficients. Regression coefficient of x on y and regression coefficient of y on x.

Assignments

1. Calculate Karl Pearson's Correlation Coefficient from the following data

X	43	44	46	40	44	42	45	42	40	42	57	48
Y	29	31	19	18	19	27	27	29	41	30	26	10

2. Compute Karl Pearson's Coefficient of Correlation from the following

Marks in Accountancy	64	56	80	45	30	60	70	20
Marks in Statistics	60	40	70	48	20	52	80	50

3. Calculate Spearman's Coefficient of Correlation from the following data

X	53	98	95	81	75	61	59	55
Y	47	25	32	37	30	40	39	45

4. Calculate the coefficient of correlation from the following data by the Spearman's Rank Differences method.

Price of Tea (Rs)	75	88	95	90	60	80	81	50
Price of Coffee (Rs)	120	134	150	115	110	140	142	100

5. Calculate the Pearson Ian Correlation Coefficient between income and weight from the following data. Also comment on the result?

Income (Rs)	100	200	300	400	500	600
Weight (Lbs.)	120	130	140	150	160	170

6. Find the Karl Pearson's Coefficient of Correlation between the following two variables. Comment on the result through the probable error?

X	06	08	12	15	18	20	24	28	31
Y	10	12	15	15	18	25	22	26	28

7. The following data relate to the age of husbands and wives. Obtain the two regression equations and determine the most likely age of husband when the age of wife is 25 years.

X	25	28	30	32	35	36	38	39	42	55
Y	20	26	29	30	25	18	26	35	35	46

8. The following table shows the exports of raw cotton and the imports of manufactured goods into India for seven years

Exports (in Crores of Rs)	42	44	58	55	89	98	60
Imports (in Crores of Rs)	56	49	53	58	67	76	58

Obtain the two regression equations and estimate the imports when exports in particular year were to the value of Rs 70 crores?

9. The following table gives the results of capital employed and profits earned by a firm in 10 successive years.

Particulars	Mean	Standard Deviation
Capital Employed (Rs Thousands)	Rs. 55	Rs.28.7
Profit Earned (Rs Thousands)	Rs. 13	Rs.8.5

Coefficient of correlation +0.96

a. Obtain the two regression equations

b. Estimate the amount of profit to be earned, if capital employed is Rs 50,000?

c. Estimate the amount of capital to be employed, if profit earned is Rs 20,000?

10. From the following data, obtain the two regression equations

Sales	91	97	108	121	67	124	51	73	111	57
Purchases	71	75	69	97	70	91	39	61	80	47

Also find correlation coefficient between sales and purchases?

11. If the regression equations between the variables x and y are $x + 6y - 6 = 0$ and $3x + 2y = 10$, find the correlation coefficient and means of the variables.

12. From the following data, find the most likely value of y when x=45. Given $r = 0.58$

	X	Y
Mean	53	142
S.D	130	165

References

1. Gujarathi , D.&Sangeetha, N. (2007). Basic Econometrics (4thed) New Delhi: McGraw Hill

2. Koutsoyianis, A. (1977). Theory of Econometrics (2nded). London. The Macmillian Press ltd

Suggested Readings

1. Anderson, D., D.Sweeney and T.Williams (2013): "Statistics for Business and Economics", Cengage Learning : New Delhi.

2. Goon, A.M. , Gupta and Das Gupta B (2002): Fundamentals of Statistics (Vol I), World Press.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

4 BLOCK

Probability

UNIT 1

Probability

Learning Outcomes

After completing this unit, the learner will be able to:

- understand the fundamental probability concepts
- interpret and apply probability definitions
- know about permutations and combinations

Background

To calculate probabilities effectively, a basic understanding of set theory, permutations, and combinations is crucial, particularly in the classical approach, which assumes all outcomes are equally likely. The empirical approach relies on familiarity with data collection and interpreting frequencies. Rooted in logical reasoning, the axiomatic approach requires a grasp of mathematical definitions and rules. Knowledge of ratios, independence, and updating probabilities based on new information is essential for understanding conditional probability.

Keywords

Set Theory, Permutations and Combinations, Probability

Discussion

4.1.1 Concept of Probability Distribution

Consider the case of tossing a coin. Nobody knows the result is head or tail. But it is certain that a head or tail will occur. In a similar way, if a die is thrown, we may get any of the faces 1, 2, 3, 4, 5 and 6. But nobody knows which one will actually occur. Experiment of this type where the outcome cannot be predicted are called random experiment. The theory of probability analyses the result obtained by such experiments.

The word probability or chance is used commonly in day-to-day life. For example, the chance of winning a game before the start of a game are equal. It is likely that Mr. Ram may not come for taking the class today. We often say that it is very probable that it will rain tomorrow. All these terms, chance, probable etc. convey the same meaning. In such of these cases we talk about chance or probability which is taken to be a quantitative measure of certainty.

In statistics, probability serves as a fundamental concept, offering a quantitative measure of the uncertainty linked to events in a random experiment. To express this uncertainty, we assign a probability value ranging from 0 to 1. A value of 1 represents complete certainty that the event will occur, while a value of 0 indicates certainty that it will not. For example, if the probability is 1/4, we interpret it as a 25% chance of occurrence and a 75% chance of non-occurrence.

This numerical assignment enables us to quantify and convey our expectations about the likelihood of different outcomes.

Expressing probability numerically and understanding its implications have practical applications across various fields, such as risk assessment and decision-making. From evaluating the chances of success in a business venture to predicting outcomes in games of chance or making informed choices under uncertain conditions, probability and its numerical representation provide a crucial tool for quantifying and managing uncertainty in a wide range of scenarios.

4.1.2 Definitions of Various Terms used in Probability

In this section, we will define and explain the various terms which are used in the definition of probability.

a. Trail and event

Consider an experiment of tossing a coin. Here tossing a coin is a trail and getting a head or tail is an event.

b. Exclusive events

The total number of possible outcomes in any trial is known as exhaustive events. For example, in tossing a coin the possible outcomes are a head and a tail. Hence, we have two exclusive events in throwing a coin.

c. Mutually exclusive events

Two events are called Mutually exclusive when the occurrence of one affects the occurrence of the other. In other words, A and B are Mutually exclusive events and if A happens then B will not happens and vice versa.

For example, in throwing a die all 6 faces numbered 1 to 6 are mutually exclusive since if anyone of these faces comes, the possibility of others, in the same trial, is ruled out.

In tossing a coin the event of head or tail are Mutually exclusive events.

d. Equally Likely

Two events are said to be equally likely if one of them cannot be expected in preference to the other.

For example, in tossing a coin head or tail are equally likely events.

e. Independent Events

Two events are said to be independent when the actual happening of one does not influence in any way happening of the other.

f. Probability

Probability of happening an event E $= \frac{\text{Number of Favourable cases}}{\text{Total Number of Cases}}$

Probability of happening of an event is p and probability of not happening of an event is q and $p + q = 1$.

Probability values are always assigned on a scale from 0 to 1. A probability near zero indicates an event is unlikely to occur; a probability near 1 indicates an event is almost certain to occur. Other probabilities between 0 and 1 represent degrees of likelihood that an event will occur.

For example, if we consider the event 'rain tomorrow,' we understand that when the weather report indicates 'a near-zero probability of rain,' it means almost no chance of rain. however, if a .90 probability of rain is reported, we know that rain is likely to occur. A .50 probability indicates that rain is just as likely to occur as not.

In discussing probability, we deal with random experiments,

g. Random Experiment

Random experiments can be defined as experiments that can be performed many times under the same conditions and their outcome cannot be predicted with complete certainty.

For example, tossing a coin and throwing a die are random experiments.

h. Sample Space

A sample space can be defined as the list of all possible outcomes of a random experiment.

4.1.3 Concept of Sets

The concept of sets is a fundamental aspect of mathematics, representing collections of distinct elements grouped together. Sets can include any type of object, such as numbers, alphabets, or even other sets, and they provide a foundational framework for understanding various mathematical structures and operations. A well-defined collection of objects or elements is a set. A 'well-defined collection' refers to a set where its members or elements are clearly defined and can be easily determined.

For example, consider the set of 'Even Natural Numbers.' This set is well-defined because it includes only those natural numbers that are divisible by 2 without leaving a remainder. Every member of this set is clearly determined: {2, 4, 6, 8, ...}.

On the other hand, a term like 'Tall People' is not a well-defined collection by itself. What exactly constitutes 'tall' is subjective and can vary from person to person, culture to culture. Hence, the set of 'Tall People' lacks a clear and universally accepted definition, making it not well-defined in a mathematical sense.

A set is an unordered collection of distinct objects, called elements or members of the set. We write $a \in A$ to denote that a is an element of the set A. The notation $a \notin A$ denotes that a is not an element of the set A. It is common for sets to be denoted using uppercase letters, while lowercase letters are typically used to denote elements of sets.

There are several ways to describe a set. One way is roster form, and another way is set builder form. In roster form, all the elements of the set are listed, separated by commas and enclosed between curly braces { }. For example, the notation {a, b, c, d} represents the set with the four elements $a, b, c, and\ d$.

The set of positive integers less than 1000 can be denoted by

{1, 2, 3…. 999}.

In Set-builder form, elements are shown or represented in statements expressing relation among elements. For example, the set A of all even positive integers less than 100 can be written as $A = \{x \mid x$ is an even positive integer less than 100}.

4.1.3.1 Set Operations

Operations on sets encompass various procedures that can be applied to create new sets or modify existing ones. Common set operations include union, intersection, difference, and complement. These operations allow us to manipulate sets to extract meaningful information and insights, providing valuable tools in fields ranging from discrete mathematics to data analysis.

A. Union of Sets

If A and B are two sets, then $A \cup B$ is the set consists of all the elements that belong to either A or B or both.

An element x belongs to the union of the sets A and B if and only if x belongs to A or x belongs to B.

i.e., $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$

Also, we can say $x \in A \cup B \Rightarrow x \in A \text{ or } x \in B$

$x \notin A \cup B \Rightarrow x \notin A \text{ and } x \notin B$

B. Intersection of sets

If A and B are two sets, then $A \cap B$ is the set consists of all the elements that belongs to both A and B.

An element x belongs to the intersection of the sets A and B if and only if x belongs to A and x belongs to B.

i.e. $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$

Also, we can say, $x \in A \cap B \Rightarrow x \in A \text{ and } x \in B$

$x \notin A \cap B \Rightarrow x \notin A \text{ or } x \notin B$

For example, $A = \{a, b, c\}$ and $B = \{a, e, f\}$ then $A \cup B = \{a, b, c, e, f\}$ and

$A \cap B = \{a\}$.

Two sets are called disjoint set if their intersection is the empty set.

The shaded region is $A \cap B$

C. Difference of sets

If A and B are two sets, then $A - B$ is the set consists of all the elements that belong to A, but are not in B. The difference of A and B is also called the complement of B with respect to A.

An element x belongs to the difference of the sets A and B if and only if x belongs to A and x does not belongs to B.

i.e. $A - B = \{x \mid x \in A \text{ and } x \notin B\}$.

For example, $A = \{a, b, c\}$ and $B = \{a, e, f\}$ then $A - B = \{b, c\}$

Note: If A and B are two sets, then,

$A = (A - B) \cup (A \cap B)$, $\quad B = (B - A) \cup (A \cap B)$

D. Complement of a set

The complement of a set is the set of all elements in a universal set that are not in the given set.

Let us denote the universal set as U, and the given set as A. The complement of set A, denoted as, A' or $\bar{A}$, consists of all elements that belong to U but do not belong to A.

An element x belongs to $\bar{A}$ if and only if $x \notin A$. i.e. $\bar{A} = \{x \in U \mid x \notin A\}$

For example, Consider the set $U = \{1,2,3,4,5,6\}$ as universe set and

$A = \{1,2,3\}$. Then the complement of $A = \{4,5,6\}$.

4.1.4 Counting Rules - Permutations and Combinations

Identifying and counting experimental outcomes is essential for accurately assigning probabilities. We now discuss three useful counting rules.

1. Counting Rule for Multiple-Step Experiments

If an experiment can be described as a sequence of k steps with n_1 possible outcomes on the first step, n_2 possible outcomes on the second step, and so on, then the total number of experimental outcomes is given by $(n_1)\,(n_2)\ldots(n_k)$.

2. Counting Rule for Combinations

A combination is a selection of objects from a group where the order is not relevant. An unordered selection of r objects from a set of n objects is called a combination.

For example, if a committee is being selected and there will be a president, vice president, and treasurer, then order matters so it is a permutation problem. But if a committee of three people is being formed without specific roles, then order does not matter and it is a combination problem.

The number of selections of n distinct objects taking r at a time is the combination denoted by $\mathrm{nC_r}$ or $C(n,r)$

$$\mathbf{nC_r} = \frac{n!}{r!\,(n-r)!}$$

3. Counting Rule for Permutations

Permutation is the arrangement of a set of elements in a particular order. From a collection of n distinct objects, any linear arrangement of these objects is called a permutation of the collection.

The number of permutations of n things taken r $(r \leq n)$ at a time is denoted by

$$nP_r \text{ or } P(n,r) = \frac{n!}{(n-r)!}$$

4.1.5 Approaches to Probability

Here, we will be discussing Classical, Empirical, and axiomatic approaches to probability.

4.1.5.1 Classical Approach

The Classical Approach to probability assumes equally likely outcomes. If an event can occur in h different ways out of a total number of n possible ways, all of which are equally likely, then the probability of the event is h/n.

For example, consider a fair six-sided die. If we want to find the probability of rolling a 4, there is only one way to roll a 4, with six possible outcomes. So, according to the Classical Approach, the probability of rolling a 4 is 1/6.

4.1.5.2 Empirical Approach / Frequency Approach

The Frequency Approach, or the Empirical Approach, relies on observed frequencies in real-world experiments. If after n repetitions of an experiment, where n is very large, an event is observed to occur in h of these, then the probability of the event is h/n. This is also called the *empirical probability* of the event.

Continuing with the die example, we might roll the die 600 times (n) in a Frequency Approach and observe that a 4 comes up 100 times. The estimated probability of rolling a 4 is then 100/600=1/6, coincidentally the same as the Classical probability. The larger the number of repetitions (n), the closer the observed frequency-based probability tends to get to the theoretical probability from the Classical Approach.

4.1.5.3 The Axiomatic Approach

The axiomatic approach of probability contains three axioms.

1. Axiom of Non-negativity

The first axiom of probability, known as the axiom of non-negativity, states that for any event A in the sample space S, the probability assigned to A, denoted as P(A), must be a non-negative real number. In mathematical terms, this means that $P(A) \geq 0$ for all events A. This axiom ensures that probabilities are always non-negative, reflecting the intuitive notion that the likelihood of any event occurring cannot be negative. Whether it is the probability of rolling a specific number on a fair die or the chance of rain on a given day, the probability values must be non-negative.

2. Axiom of Normalization

The second axiom, often referred to as the axiom of normalization or the sum rule, asserts that the probability of the entire sample space S equals 1. Symbolically, this can be expressed as $P(S) = 1$. In other words, the sum of probabilities for all possible outcomes or events in the sample space is unity. This axiom ensures that the total probability mass in the sample space is accounted for, aligning with the idea that, under any circumstance, something in the sample space must occur.

3. Axiom of Additivity

The third axiom, the axiom of additivity, addresses the probabilities of combinations of events. It states that for any sequence of mutually exclusive events $\{A_1, A_2, ..., An\}$, the probability of the union of these events is equal to the sum of their probabilities. Mathematically, this is expressed as $P(A_1 \cup A_2 \cup ... \cup An) = P(A_1) + P(A_2) + ... + P(An)$. The additivity axiom handles scenarios where events are not mutually exclusive by adjusting for their potential overlap. It ensures a coherent and consistent way to calculate probabilities for complex situations involving multiple events.

Illustration 4.1.1

A uniform die is thrown at random. Find the probability that the number on it is

(i) 5, (ii) greater than 4, (iii) even.

Solution

Since the dice can fall with any one of the faces 1, 2, 3, 4, 5, and 6, the exhaustive number of cases is 6.

The number of cases favourable to the event of getting '5' is only 1.

∴ Required probability = 1/6.

The number of cases favourable to the event of getting a number greater than 4 is 2, viz., 5 and 6. ∴ Required probability $= \frac{2}{6} = \frac{1}{3}$

Favourable cases for getting an even number are 2, 4 and 6, i.e., 3 in all.

∴ Required probability $= \frac{3}{6} = \frac{1}{2}$

Illustration 4.1.2

If six dice are rolled, find probability that all show different faces?

Solution

In a random roll of six dice, the exhaustive number of cases is $n(S) = 6^6$.

Define the event E : All the six dice show different faces.

We can get any one of the six faces 1, 2, 3, 4, 5, 6, on the first dice. For the happening of E, the second die must show any one of the remaining 5 faces, the third die must show any one of the remaining 4 faces, and so on, the 6th die must show the remaining last face.

Hence, by the principle of counting, the number of cases favourable to the happening of E are

$$n(E) = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 6!$$

$$\therefore\ P(E) = \frac{n(E)}{n(S)}$$

$$= 6!/6^6$$

Illustration 4.1.3

A bag contains 20 tickets marked with numbers 1 to 20. One ticket is drawn at random. Find the probability that it will be a multiple of (i) 2 or 5, (ii) 3 or 5.

Solution

One ticket can be drawn out of 20 tickets in $20C_1 = 20$ ways, which determine the exhaustive number of cases.

i. The number of cases favourable to getting the ticket number which is

- a multiple of 2 are 2, 4, 6, 8, 10, 12, 14, 16, 18, 20 i.e., 10 cases.
- a multiple of 5 are 5, 10, 15, 20 i.e., 4 cases. Of these, two cases viz., 10 and 20 are duplicated. Hence the number of distinct cases favourable to getting a number which is a multiple of 2 or 5 are 10 + 4 – 2 = 12. ∴ Required probability = 12 ÷ 20 = 3 ÷ 5 = 0·6.

ii. The cases favourable to getting a multiple of 3 are 3, 6, 9, 12, 15, 18 i.e., 6 cases in all and getting a multiple of 5 are 5, 10, 15, 20 i.e., 4 cases in all. Of these, one case viz., 15 is duplicated. Hence, the number of distinct cases favourable to getting a multiple of 3 or 5 is 6 + 4 – 1 = 9. ∴ Required probability $= \frac{9}{20}$ = 0·4

Illustration 4.1.4

An urn contains 8 white and 3 red balls. If two balls are drawn at random, find the probability that (i) both are white (ii) both are red (iii) one is of each colour.

Solution

Total number of balls in the urn is $8 + 3 = 11$. Since 2 balls can be drawn out of 11 balls in $11C_2$ ways,

Exhaustive number of cases $= 11C_2 = \frac{11 \times 10}{2} = 55$

i. If both the drawn balls are white, they must be selected out of the 8 white balls and this can be done in $8C_2 = \frac{8\times7}{2} = 28$ ways. Probability that both the balls are white = 28

ii. If both the drawn balls are red, they must be drawn out of the 3 red balls and this can be done in $3C_2 = 3$ ways. Hence, the probability that both the drawn balls are red = 3

iii. The number of favourable cases for drawing one white ball and one red ball is $8C_1 \times 3C_1 = 8 \times 3 = 24$

∴ Probability that one ball is white and other is red = 24

Illustration 4.1.5

A student is to answer seven out of 10 questions on an examination. In how many ways can he make his selection if (a) there are no restrictions? (b) he must answer the first two questions? (c) he must answer at least four of the first six questions?

Solution

Number of ways of selection of 7 questions out of 10

$$10C_7 = 10C_3 = \frac{10 \times 9 \times 8}{3 \times 2 \times 1} = 120$$

i.e., in 120 ways.

After selecting first two questions, he must select 5 questions from the remaining 8 questions. This can be done in $8C_5 = 8C_5 = \frac{8 \times 7 \times 6}{3 \times 2 \times 1} = 56$

i.e., 56 ways.

He must answer four of the first 6 questions and 3 from the remaining 4 questions in $(6C_4 \times 4C_3)$ ways, $6C_4 \times 4C_3 = 15 \times 4 = 60$

Four of the first 5 questions and 2 from the remaining 4 questions in $(6C_5 \times 4C_2)$ ways, six of the first 6 questions and 1 from the remaining 4 questions in

$(6C_6 \times 4C_1)$ ways. $6C_5 \times 4C_2 = 6 \times 6 = 36$

$$\binom{6}{5} \times \binom{4}{2} = 6 \times 6 = 36$$

Therefore, number of ways of selecting at least four of the first six questions

$= 6C_4 \times 4C_3 + 6C_5 \times 4C_2 + 6C_6 \times 4C_1 = 100$ ways.

Illustration 4.1.6

An urn contain 15 balls 8 of which are red and 7 are black. In how many ways can 5 balls be chosen so that a) all five are red b)) all five are black c) 2 are red and 3 are black d) 3 are red and 2 are black.

Solution

Five red balls can be selected from 8 red balls,

$$8C_5 = \frac{8 \times 7 \times 6}{3 \times 2 \times 1} = 56$$

Five black balls can be selected from 7 black balls,

$$7C_5 = 7C_2 = \frac{7 \times 6}{2 \times 1} = 21$$

i.e., 21 ways.

Two red balls can be selected from 8 red balls

$$8C_2 = \frac{8 \times 7}{2 \times 1} = 28$$

i.e., 28 ways

and three black balls can be selected from 7 black balls

$$7C_3 = \frac{7 \times 6 \times 5}{3 \times 2 \times 1} = 35$$

i.e., in 35 ways.

By the rule of product, the selection can be made in 28×35= 980 ways.

Three red balls can be selected from 8 red balls

$$8C_3 = \frac{8 \times 7 \times 6}{3 \times 2 \times 1} = 56$$

i.e., 56 ways

and two black balls can be selected from 7 black balls

$$7C_2 = \frac{7 \times 6}{2 \times 1} = 21$$

i.e., in 21 ways.

By the rule of product the selection can be made in 56×21= 1176 ways.

Summarised Overview

A set is an unordered collection of distinct objects, called elements or members of the set. The number of selections of n distinct objects taking r at a time is the combination denoted by nC_r or $C(n, r)$. Permutation is the arrangement of a set of elements in a particular order. Probability is a measure of the likelihood of an event occurring, expressed as a value between 0 and 1.

In the Classical Approach to probability, outcomes are assumed to be equally likely, and the probability of an event is calculated as the ratio of favourable outcomes to the total possible outcomes. The Empirical Approach to probability involves estimating the likelihood of an event based on observed frequencies from real-world experiments or data. The Axiomatic Approach formulates probability theory using a set of axioms or fundamental principles, providing a rigorous mathematical foundation for probability.

Assignments

1. How many ways can we arrange the letter in the word COMPUTER. Find the number of permutations of size 2. If reparations are allowed, find the number of possible 12 letter sequence.

2. In how many ways can 12 different books be distributed among four children so that a) each child gets three books? b) the two oldest children get four books each and the two youngest get two books each?

3. The letters of the word 'article' are arranged at random. Find the probability that the vowels may occupy the even places**.**

4. If a pair of dice is thrown, find the probability that the sum of the digits on them is neither 7 nor 11

5. A bag contains 8 black, 3 red and 9 white balls. If 3 balls are drawn at random, find the probability that (a) all are black, (b) 2 are black and 1 is white, (c) 1 is of each colour, (d) the balls are drawn in the order black, red and white, (e) None is red

6. Five digited numbers are formed from the digits 1, 2, 3, 4, 5. Find the chance that the number formed is greater than 23000.

References

1. Gujarathi , D.&Sangeetha, N. (2007). *Basic Econometrics* (4thed) New Delhi: McGraw Hill

2. Koutsoyianis, A. (1977). *Theory of Econometrics* (2nded). London. The Macmillian Press ltd

Suggested Readings

1. Anderson, D., D.Sweeney and T. Williams (2013): "*Statistics for Business and Economics*", Cengage Learning : New Delhi.

2. Goon, A.M. , Gupta and Das Gupta B (2002): *Fundamentals of Statistics* (Vol I), World Press.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

2

UNIT

Probability and Random Variables

Learning Outcomes

After completing this unit, the learner will be able to:

- apply addition and multiplication laws to calculate probabilities of events
- analyse conditional probability and apply Bayes' theorem
- use discrete distributions to model and analyse event outcomes
- examine relationships between two random variables

Background

To understand the concepts of two random variables, joint distributions, and their expectations, it is essential to have a strong grasp of foundational probability concepts. Addition and multiplication laws form the basis for calculating the probabilities of unions and intersections of events. Conditional probability helps to quantify the likelihood of an event given that another event has occurred, while Bayes' theorem enables the updating of probabilities based on new information. A random variable associates numerical values to outcomes of a random experiment, and its probability distribution describes how probabilities are distributed across these values. The concept of mathematical expectation (or expected value) generalises the mean of a probability distribution, while moments (such as variance, skewness) describe its shape and spread. For two random variables, their joint probability distribution models the simultaneous behaviour of both, and their joint expectation provides insights into their relationship. These concepts are critical for understanding dependencies, correlations, and making predictions in applied contexts.

Keywords

Conditional probability, Random variables, joint, Random variables

4.2.1 Addition Law of Probability

Let S be the sample space of a given experiment. Let A and B be two events of S. $A \cup B$ denote the event that event A or event B (or both) occur when the experiment is performed. $A \cap B$ denotes the event that both A and B occur together.

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

This rule can be extended to three or more events, for example:

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

If two events A and B are mutually exclusive then

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

4.2.2 Multiplication Law of probability

If A and B are independent events then

$$P(A \cap B) = P(A) \times P(B)$$

i.e.. The probability of independent events A and B occurring is the product of the probabilities of the events occurring separately.

4.2.3 Conditional probability

Suppose a bag contains 6 balls, 3 red and 3 white. Two balls are chosen (without replacement) at random, one after the other. Consider the two events,

R is the event that the first ball chosen is red ,

W the event that the second ball chosen is white.

We easily find $P(R) = \frac{3}{6} = \frac{1}{2}$. If the first ball chosen is red then the bag subsequently contains 2 red balls and 3 white. In this case $P(W) = \frac{3}{5}$. However, if the first ball chosen is white then the bag subsequently contains 3 red balls and 2 white. In this case $P(W) = \frac{2}{5}$. i.e, the probability that W occurs is clearly dependent upon whether or not the event R has occurred. The probability of W occurring is conditional on the occurrence or otherwise of R. The conditional probability of an event B occurring given that event A has occurred is written $P(B|A)$. In this particular example $P(W|R) = \frac{3}{5}$ and $P(W|R') = \frac{2}{5}$.

Which shows that the probability that W occurs is clearly dependent upon whether or not the event R has occurred. The probability of W occurring is conditional on the occurrence or otherwise of R. The conditional probability of an event B occurring given that event A has occurred is written $P(B|A)$.

The conditional probability $P(B|A)$ is defined as

$$P(B|A) = \frac{number\ of\ outcomes\ in\ A\cap B}{number\ of\ outcomes\ in\ A} = \frac{P(A\cap B)}{P(A)}$$

Or

$$P(A \cap B) = P(B|A) \times P(A)$$

4.2.4 Baye's Theorem

Let $A_1, A_2, \ldots A_n$ be n exclusive and exhaustive events with $P(A_i) \neq 0$ for $i = 1,2, \ldots n$. Let B be an event such that $P(B) > 0$, Then

$$P(A_I|B) = \frac{P(A_i).\ P(B|A_i)}{\sum_{i=1}^{n} P(A_i).\ P(B|A_i)}$$

Illustration 4.2.1

A person is known to hit the target in 3 out of 4 shots, where as another person is known to hit the target in 2 out of 3 shots. Find the probability of the targets being hit at all when they both persons try?

Solution

Probability of the first person hit the target = $P(A) = \frac{3}{4}$

Probability of the second person hit the target = $P(B) = \frac{2}{3}$

The two events are not mutually exclusive, since both persons hit the same target.

∴ Required probability P(A or B) = $P(A) + P(B) - P(A \cap B)$

$$= \left(\frac{3}{4} + \frac{2}{3}\right) - \left(\frac{3}{4} \times \frac{2}{3}\right) \quad since\ A\ and\ B\ are\ independent$$

$$= \frac{17}{12} - \frac{6}{12}$$

$$= \frac{11}{12}$$

Illustration 4.2.2

A bag contains 20 balls,3 are coloured red, 6 are coloured green, 4 are coloured blue, 2 are coloured white and 5 are coloured yellow. One ball is selected at random. Find the probabilities of the following events. (a) the ball is either red or green (b) the ball is not blue (c) the ball is either red or white or blue.

Solution

P [getting Red ball] $= \frac{3}{20}$ P[getting green ball] $= \frac{6}{20}$ P[getting blue ball] $= \frac{4}{20}$

P [getting white ball] $= \frac{2}{20}$

P [the ball is either red or green] $= \frac{3}{20} + \frac{6}{20} = \frac{9}{20}$

P [the ball is blue] $= \frac{4}{20}$, P [the ball is not blue] $= 1 - \frac{4}{20} = \frac{16}{20} = \frac{4}{5}$

P [the ball is either red or white or blue] $= \frac{3}{20} + \frac{2}{20} + \frac{4}{20} = \frac{9}{20}$.

Illustration 4.2.3

If from a pack of cards a single card is drawn. What is the probability that it is either spade or king?

Solution

$P[A] = P[Spade\ card] = 13/52$

$P[B] = P[King\ \ card] = 4/52$

$P[\text{either spade or king}] = P(A) + P(B) - P(A \cap B)$

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

$$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$$

Illustration 4.2.4

The probability that machine A will be performing an usual function in 5 years time is 1/4 while the probability that the machine B will still be operating usefully at the end of the same period is 1/3 . Find the probability that both machines will be performing an usual function

Solution

P[machine A operating usually] $= \frac{1}{4}$

P[machine B operating usually] $= \frac{1}{3}$

P[both machines will be performing usual function] $= P(A) \times P[B] = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}$

Illustration 4.2.5

A bag contain 8 white and 10 black balls. Two balls are drawn in succession. What is the probability that first is white and the second is black?

Solution

Total number of balls = 8+10 = 18

P[drawing a white ball] $= \frac{8}{18} = \frac{4}{9}$

After drawing a white ball there are 17 balls in the bag .

P[drawing a black ball] = 10/17

P[drawing first white and the second black ball] $== \frac{4}{9} \times \frac{10}{17} = \frac{40}{153}$

Illustration 4.2.6

Let 5 men out of 100 and 25 women out of 1000 are colour blind. A colour blind person is chosen at random. What is the probability of his being male. (Assume that males and female are in equal proportions).

Solution

Let M denote a person is Male. Let F denotes a person is Female. Let C denote a person is colour blind.

Given $P(M) = \frac{1}{2}, \quad P(F) = \frac{1}{2}$

$$P(C|M) = \frac{5}{100}, \qquad P(C|F) == \frac{25}{1000}$$

$$P(C|M) = \frac{P(C|M).P(M)}{P(C|M)P(M) + P(C|F)P(F)}$$

$$= \frac{\frac{5}{100}\cdot\frac{1}{2}}{\frac{5}{100}\cdot\frac{1}{2} + \frac{25}{1000}\cdot\frac{1}{2}}$$

$$= \frac{\frac{1}{40}}{\frac{1}{40} + \frac{1}{80}} = \frac{2}{3}$$

Illustration 4.2.7

An urn contain 10 W, 3B balls while another urn contain 3 W, 5 B balls. Two balls are drawn from the first urn and put into the second urn. Then a ball is drawn from the latter. What is the probability that it is a white ball

Solution

Two balls drawn from the 1^{st} urn may be

both white (event A_1)

both black (event A_2)

1 W, 1 B (event A_3)

$$\therefore P(A_1) = \frac{10C_2}{13C_2} = \frac{48}{78} = \frac{15}{26}$$

$$P(A_2) = \frac{3C_2}{13C_2} = \frac{1}{26}$$

$$P(A_3) = \frac{10C_1 \times 3C_1}{13C_2} = \frac{10}{26}$$

After the balls are transformed from 1^{st} urn to 2^{nd} urn, the 2^{nd} urn will contain

i) 5W, 5B ii) 3 W, 7 B iii) 4W, 6B

Let B be the event of drawing a white ball from the 2^{nd} urn.

Now $P(B|A_1) = \frac{5C_1}{10\,C_1} = \frac{5}{10}$

$$P(B|A_2) = \frac{3C_1}{10C_1} = \frac{3}{10}$$

$$P(B|A_3) = \frac{4C_1}{10C_1} = \frac{4}{10}$$

$$\therefore P(B) = \sum_{i=1}^{3} P(B|A_i)\,P(A_i)$$

$$= \frac{5}{10}\cdot\frac{15}{26} + \frac{3}{10}\cdot\frac{1}{26} + \frac{4}{10}\cdot\frac{10}{26} = \frac{59}{130}$$

4.2.4 Mathematical Expectation

If X is a random variable taking the values $x_1, x_2 \ldots . x_n$ with corresponding probability $f_1, f_2 \ldots f_n$ then mathematical expectation od X is denoted by $E(X)$ and is given by

$E(X) = \sum_x x\, f(x)$ (for discrete random variables

$= \int_{-\infty}^{\infty} x\, f(x)dx$ (for continuous random variable)

$E(X^2) = \sum_x x^2\, f(x)$ (for discrete random variables

$= \int_{-\infty}^{\infty} x^2\, f(x)dx$ (for continuous random variable)

Mean and variance of a distribution can represent in terms of expectation

Mean of a distribution is $E(X)$ and variance of a distribution is

$V(X) = E(X^2) - \left(E(X)\right)^2$

4.2.5 Moments (r^{th} Moment About Origine)

Consider the discrete random variable X with probability density function f(x) then the r^{th} moment about origine of the probability distribution is defined as

$E(X^r) = \sum x^r\, f(x)$

It is denoted by $\mu_r' = \sum x^r\, f(x)$

Put $r = 1,\ \mu_1' = \sum x\, f(x) = E(X)$

Put $r = 2,\ \mu_2' = \sum x^2\, f(x) = E(X^2)$

Mean = E(X) and Variance = $E(X^2) - \left(E(X)\right)^2$

For contineous random variables $\mu_r' = \int x^r\, f(x)\, dx$

r^{th} Moment about mean

$\mathrm{E}\left(\mathrm{X} - \mathrm{E}(\mathrm{X})\right)^{\mathrm{r}} = E(X - \bar{x})^r = \Sigma(X - \bar{x})^r\, f(x)$ for discrete random variables

$= \int (X - \bar{x})^r\, f(x)dx$ for continuous random variables

Is the r^{th} moment about mean of a random variable. It is denoted by μ_r

Put $r = 1,\quad \mu_1 = 0$ is the mean.

Put $r = 2,\quad \mu_2 = \sum (X - \bar{x})^2\, f(x)$ = variance .

Variance $\mu_2 = \mu_2' - (\mu_1')^2$

4.2.6 Two Dimensional Random Variable

On studying random variables and their probability distributions we restricted our selves to one dimensional sample space. The outcomes of the experiment are the values assumed by a single random variable. However there will be situations where the outcome of several random variables simultaneously.

For example, income X and consumption expenditure Y for individuals in an economy. These two variables are often correlated because higher income typically leads to higher consumption. This gives a Two Dimensional sample space consisting of outcomes (X,Y). So here Two Dimensional Random variables X and Y are involved simultaneously.

4.2.7 Joint Probability Distribution

If X and Y are two random variables, The probability distribution of their simultaneous occurrences can be represented by a function $f(x,y)$, for any pair of values (x,y) within the range of the random variables X and Y. This function is known as probability distribution of X and Y.

$$f(x,y) = P[X = x, Y = y]$$

If the possible vales of X and Y are finite or countably infinite then (X,Y) is called a Two Dimensional discrete Random variable.

The function P(x,y) is the joint probability mass function of a discrete Random variable (X,Y) if

$P(X = x_i, Y = y_i) \geq 0$ where $i = 1,2,3 \dots, \quad j = 1,2,3 \dots.$

$$\Sigma_i \Sigma_j P(X = x_i, Y = y_i) = 1$$

If the possible vales of X and Y assumes all the vales in a specific region in XY plane then (X,Y) is called a Two Dimensional continuous random variable.

If (X,Y) is a two-dimensional continuous random variable such that

$P\left\{x - \frac{dx}{2} \leq X \leq x + \frac{dx}{2}, y - \frac{dy}{2} \leq Y \leq y + \frac{dy}{2}\right\} = f(x,y)$ then $f(x,y)$ is called the joint probability density function of (X,Y) provided $f(x,y)$ satisfies the following conditions.

$f(x,y) \geq 0$ for all $(x,y) \in R$ where R is the range space.

$$\iint_R f(x,y)dx\, dy = 1$$

4.2.8 Marginal Distribution

Let (X,Y) be two dimensional random variable

Discrete Case

The marginal distribution for X alone is given by

$$P[X = x_i] = \sum_j P[X = x_i, Y = y_j] = p_{i.}$$

The marginal distribution for Y alone is given by

$$P[Y = y_j] = \sum_i P[X = x_i, Y = y_j] = p_{.j}$$

Continuous Case

The marginal distribution for X alone is given by

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

The marginal distribution for Y alone is given by

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

4.2.9 Conditional Probability Distribution

Discrete Case

Let $P[X = x_i,\ Y = y_i]$ be the joint probability function of a two dimensional random variable (X,Y). The conditional probability function of X given $Y = y_j$

$$P[X = x_i|Y = y_j] = \frac{P[X = x_i \cap Y = y_j]}{P[Y = y_j]} = \frac{p_{ij}}{p_{.j}}$$

The conditional probability function of Y given $X = x_i$ is defined by

$$P\left[Y = y_j \middle| X = x_i\right] = \frac{P[X = x_i \cap Y = y_j]}{P[X = x_i]} = \frac{p_{ij}}{p_{i.}}$$

Continuous Case

Let (X, Y) be the two dimensional continuous random variables with joint probability density function $f(x, y)$. Then the conditional density function of X given Y is

$$f(x/y) = \frac{f(x, y)}{f_Y(y)}$$

Where $f_Y(y)$ is the marginal density function of Y.

Similarly the conditional density function of Y given X is

$$f(y/x) = \frac{f(x,y)}{f_X(x)}$$

Where $f_X(x)$ is the marginal density function of X

If X and Y are two dimensional continuous random variables with joint probability density function $f(x,y)$ such that $f(x,y) = f(x).f(y)$ then X and Y are said to be independent random variables.

4.2.10 Expectation of a Function $f(x,y)$

The expectation of a function f(x,y) of a two dimensional continuous random variables with joint probability density function $f(x,y)$ is

$$E\big(f(x,y)\big) = \begin{Bmatrix} \sum_i \sum_j f(x_i,y_j)\, P[X = x_i | Y = y_j] \ (Discrete) \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_i,y_j)\ \ f_{xy}(x,y) dx\, d\,y \ (Continuous) \end{Bmatrix}$$

Illustration 4.2.8

The joint probability function (X,Y) is given by $P(X,Y) = k(2x + 3y), \ \ x = 0,1,2,$

$y = 1,2,3$ 1) Find the marginal distribution

2) Conditional probability of Y given x = 1

Solution

Y X	1	2	3	Total $P_X(x)$
0	3k	6k	9k	18k
1	5k	8k	11k	24k
2	7k	10k	13k	30k
Total $P_Y(y)$	15k	24k	33k	72k

We know that $\sum P_{ij} = 1 \ \ \Rightarrow \ \ 72k = 1 \Rightarrow k = \frac{1}{72}$

Hence joint probability function is

Y / X	1	2	3	Total $P_X(x)$
0	3/72	6/72	9/72	18/72
1	5/72	8/72	11/72	24/72
2	7/72	10/72	13/72	30/72
Total $P_Y(y)$	15/72	24/72	33/72	72/72

Marginal distribution of X

X	0	1	2
$P_X(x)$	18/72	24/72	30/72

Conditional probability of Y given x = 1

$$P[X = 0 \mid Y = 1] = \frac{\frac{3}{72}}{\frac{15}{72}} = \frac{3}{15} = \frac{1}{5}$$

$$P[X = 1/ Y = 1] = \frac{\frac{5}{72}}{\frac{15}{72}} = \frac{5}{15} = \frac{1}{3}$$

$$P[X = 2 \mid Y = 1] = \frac{\frac{7}{72}}{\frac{15}{72}} = \frac{7}{15}$$

X	0	1	2
$P[X = 0 \mid Y = 1]$	1/5	1/3	7/15

Illustration 4.2.9

Let X and Y have the following joint probability distribution

X Y	2	4
1	0.1	0.15
3	0.2	0.3
5	0.1	0.15

Solution

X Y	2	4	$P_{j\cdot}$
1	0.1	0.15	0.25
3	0.2	0.3	0.5
5	0.1	0.15	0.25
$P_{\cdot\, i}$	0.4	0.6	1

Marginal distribution of Y on X

Y	1	3	5
$P_Y(y)$	0.25	0.5	0.25

Marginal distribution of X on Y

X	2	4
$P_Y(y)$	0.4	0.6

$p_{1.} \times p_{.2} = 0.25 \times 0.6 = 0.15 = p_{12}$

i.e.

$p_{i.} \times p_{.j} = p_{ij}$

$\therefore$ X and Y are independent

Illustration 4.2.10

The joint probability distribution of two dimensional random variable is

$$f(x,y) = \frac{8}{9}\,xy \quad 1 < x < y < 2$$

$$= 0 \text{ otherwise}$$

Find 1) Marginal density function of X

2) Conditional probability of Y given $X = x$

Solution

$$f_X(x) = \int_x^2 \frac{8}{9}\, xy\, dy$$

$$= \frac{8}{9}x\left(\frac{y^2}{2}\right)_x^2 = \frac{4x}{9}(4 - x^2)\ , \qquad 1 \le x \le 2$$

$$f_Y(y) = \int_1^y \frac{8}{9}\, xy\, dx$$

$$= \frac{8}{9}y\left(\frac{x^2}{2}\right)_1^y = \frac{4y}{9}(y^2 - 1)\ , \qquad x \le y \le 2$$

Conditional probability

$$f(x/y) = \frac{f(x,y)}{f_Y(y)}$$

$$= \frac{\frac{8}{9}xy}{\frac{4y}{9}(y^2 - 1)},$$

$$= \frac{2x}{(y^2 - 1)}, \qquad 1 \le x \le y \le 2$$

$$f(y/x) = \frac{f(x,y)}{f_X(x)}$$

$$= \frac{\frac{8}{9}xy}{\frac{4x}{9}(4 - x^2)}$$

$$= \frac{2y}{(4 - x^2)}, \qquad 1 \le x \le y \le 2$$

Illustration 4.2.11

The joint probability distribution of X and Y is

$$f(x,y) = e^{-(x+y)}, \quad 0 \leq x, \quad y \leq \infty$$

$$= 0 \text{ otherwise}$$

Are X and Y independent?

Solution

$$f_X(x) = \int_0^{\infty} e^{-(x+y)}\, dy$$

$$= e^{-x} \int_0^{\infty} e^{-y}\, dy$$

$$= e^{-x}[-e^{-y}]_0^{\infty}$$

$$= e^{-x}[-e^{-\infty} + 1]$$

$$= e^{-x}[0 + 1]$$

$$= e^{-x}$$

$$f_Y(y) = \int_0^{\infty} e^{-(x+y)}\, dy$$

$$= e^{-y} \int_0^{\infty} e^{-x}\, dy$$

$$= e^{-y}[-e^{-x}]_0^{\infty}$$

$$= e^{-y}[-e^{-\infty} + 1]$$

$$= e^{-y}[0 + 1]$$

$$= e^{-y}$$

$$f_X(x) \times f_Y(y) = e^{-x} \times e^{-y} = e^{-(x+y)} = f(x,y)$$

Illustration 4.2.12

Let X and Y have the joint probability distribution

X Y	0	1	2
0	0.1	0.4	0.1
1	0.2	0.2	0

Find

$P(X+Y>1)$

The probability function $P(X=x)$ of the random variable.

$P[Y=1/\ X=1]$

$E(XY)$

Solution

$$P(X+Y>1) = P(1,1) + P(2,0) + P(2,1)$$

$$= 0.2 + 0.1 + 0 = 0.3$$

The probability function of X is

X	0	1	2
$P(X=x)$	0.3	0.6	0.1

$$P[Y=1/\ X=1] = \frac{P[Y=1,X=1]}{P[X=1]} = \frac{0.2}{0.6} = \frac{1}{3}$$

$$E(XY) = \sum_{x=0}^{2}\sum_{y=0}^{1} xy\ \ p(x,y)$$

$$= (1\times 1)p(1,1) + (2\times 1)p(2,1)$$

$$= 1\times 0.2 + 2\times 0$$

$$= 0.2$$

Summarised Overview

Addition law is represented as -$P(A \cup B) = P(A) + P(B) - P(A \cap B)$. In the case of Multiplication law- If A and B are independent events then

$P(A \cap B) = P(A) \times P(B)$

Baye's theorem - Let $A_1, A_2, \dots A_n$ be n exclusive and exhaustive events with $P(A_i) \neq 0$ for $i = 1,2, \dots n$. Let B be an event such that $P(B) > 0$, Then

$P(A_I|B) = \dfrac{P(A_i).\ P(B|A_i)}{\sum_{i=1}^{n} P(A_i).\ P(B|A_i)}$ If X is a random variable taking the values $x_1, x_2 \dots . x_n$ with corresponding probability $f_1, f_2 \dots f_n$ then mathematical expectation od X is denoted by $E(X)$ and is given by

$$E(X) = \sum_x f(x)$$

If X and Y are two random variables, The probability distribution of their simultaneous occurrences can be represented by a function $f(x, y)$, for any pair of values (x, y) within the range of the random variables X and Y.

Assignments

1. An urn contains 7 white and 3 red balls. Two balls are drawn together, at random, from this urn. Compute the probability that neither of them is white. Find also the probability of getting one white and one red ball. Hence compute the expected number of white balls drawn

2. An urn contain 10 white and 3 black balls. Another urn contain 3 white and 5 balck balls. 2 balls are drawn at random from the first urn and placed in the second urn and then one ball is taken at random from the latter. What is the probability that it is a white ball?

3. A person is known to hit the target in 3 out of 4 shots whereas another person is known to hit the target in 2 out of 3 shots, Find the probability of the target being hit at all when the both persons try

4. An urn contain 5 balls. 2 balls are drawn and are found to be white. What is the probability of all the balls being white?

5. A continuous random variable X has the probability density function

$$f(x,y) = \frac{1}{2}(x+1), \quad -1 < x < 1,$$

= 0 otherwise

Find the Mean and variance of X

6. The joint probability distribution of two dimensional random variable is

$$f(x,y) = 24\, y\,(1-x) \quad 0 < y < x < 1$$

$= \; 0$ otherwise

Find the Marginal density function of X. Also find E(XY)

7. The joint probability distribution of two dimensional random variable is

$$f(x,y) = \; x(1 + 3y^2) \quad 1 < x < 2{,}0 < y < 1$$

$= \; 0$ otherwise

Find 1) Marginal density function of X

2)Conditional probability of X given Y

References

1. Gujarathi , D.&Sangeetha, N. (2007). *Basic Econometrics* (4thed) New Delhi: McGraw Hill

2. Koutsoyianis, A. (1977). *Theory of Econometrics* (2nded). London .The Macmillian Press ltd

Suggested Readings

1. Anderson, D., D.Sweeney and T.Williams (2013): *"Statistics for Business and Economics"*, Cengage Learning : New Delhi.

2. Goon, A.M. , Gupta and Das Gupta B (2002): *Fundamentals of Statistics* (Vol I), World Press.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

3

UNIT

Applications of Discrete Distributions

Learning Outcomes

After completing this unit, learners will be able to:

- understand probability distributions
- analyse distribution functions
- know binomial, poisson and uniform distributions

Background

To study probability distributions, distribution functions, and specific distributions like binomial and normal, one must understand the basics of probability theory, including concepts like random variables, outcomes, and events. Familiarity with descriptive statistics is crucial for interpreting distribution parameters. Knowledge of combinatorics (e.g., combinations) helps in understanding the binomial distribution, while a basic knowledge of calculus, particularly integration, is essential for continuous distributions like the normal distribution. Additionally, a foundation in basic algebra and graphing functions is helpful for visualising and analysing distribution curves and their behaviour.

Keywords

Probability Distributions, Binomial, Poisson, Uniform Distributions

4.3.1 Random Variable

A Random Variable is a function that assigns numerical values to the outcomes of a random experiment. In other words, a random variable converts chance-based outcomes into numbers. Let us take a simple experiment, throwing a coin twice. The possible outcomes of this experiment are:

S = {HH, HT, TH, TT}, where H stands for Head, T stands for Tail. Now, define a variable X to represent the number of heads obtained in each outcome. Let us look at each outcome and the number of heads:

Event/Outcome	HH	HT	TH	TT
Value/ Number of Heads	2	1	1	0

So, X can take the values 0, 1, or 2. This variable X is called a Random Variable because it depends on the outcome of a random experiment and assigns a numerical value to each possible outcome in the sample space. Random variables are mainly classified into two types:

- **Discrete Random Variable**: A Discrete Random Variable is one that takes only specific, countable values. These values can be listed or counted, even if they extend to infinity. Such variables usually represent quantities that are counted rather than measured. For example, the number of students in a class can be 0, 1, 2, 3, and so on - these are distinct and countable values. Similarly, the number of heads obtained when flipping 3 coins can be 0, 1, 2, or 3. In both cases, the values are limited to specific outcomes and do not include fractions or decimals.

- **Continuous Random Variable**: A Continuous Random Variable is a type of variable that can assume any value within a specified range, including all possible decimal and fractional values. For instance, measurements such as the height of students (e.g., 155.5 cm, 160.2 cm) or the time taken to complete a task (e.g., 2.3 hours, 2.31 hours) are examples of continuous random variables. These variables can be measured but not counted exactly, as there are infinitely many possible values within any interval.

4.3.1.1 Discrete Random Variables

Discrete Probability Distributions (Probability Mass Function)

Let X be a discrete random variable, and suppose that the possible values that it can assume are given by $x_1, x_2, x_3, \ldots$, arranged in some order. Suppose also that these values are assumed with probabilities given by

$$P(X = x_k) = f(x_k) \; k = 1,2, \ldots \qquad (1)$$

It is convenient to introduce the *probability function*, also referred to as a *probability distribution*, given by

$$P(X = x) = f(x) \qquad (2)$$

For a discrete random variable X, the set of possible values it can assume is denoted by $x_1, x_2, x_3, \ldots$, arranged in some order. The probabilities associated with each of these values are specified by the probability mass function (PMF), denoted by $f(x_k)$, where x_k represents a particular value that X can take. The PMF describes the likelihood of X assuming each of its possible values.

For a discrete random variable X, the set of possible values it can assume is denoted by $x_1, x_2, x_3, \ldots$, arranged in some order. The probabilities associated with each of these values are specified by the probability mass function (PMF), denoted by $f(x_k)$, where x_k represents a particular value that X can take. The PMF describes the likelihood of X assuming each of its possible values.

The probability mass function is formally expressed as:

$$P(X = x_k) = f(x_k) \text{ for } k = 1, 2, \ldots \qquad (1)$$

Here, $P(X = x_k)$ represents the probability that the discrete random variable X takes on the specific value x_k, and $f(x_k)$ is the associated probability mass function for that value.

To streamline notation, it is convenient to introduce the probability function or probability distribution, denoted by $P(X = x)$, where x is a generic variable representing any of the possible values that X can assume. This probability function is defined by:

$$P(X = x) = f(x) \qquad (2)$$

In summary, the probability mass function $f(x)$ specifies the probabilities associated with individual values of a discrete random variable, and the probability distribution $P(X = x)$ generalises this to express the probability of X being any specific value x.

For $x = x_k$, this reduces to (1) while for other values of x, $f(x) = 0$.

In general, $f(x)$ is a probability function if

$$f(x) \geq 0$$

$$\Sigma f(x) = 1$$

4.3.2 Binomial Distribution

Let us assume a salesperson is attempting to close deals with potential clients. Each interaction with a client can be viewed as a trial, and the outcome of each trial is either a successful deal (success) or an unsuccessful attempt (failure). Let us denote the probability of successfully closing a deal in any single trial as p, and the probability of failure as $q = 1 - p$.

Now, suppose the salesperson conducts a series of n trials, trying to close deals with different clients independently. The objective is to understand the probability distribution of the number of successful deals (x) among these n trials. This is where the probability function for a binomial distribution comes into play.

The probability mass function (PMF) for a binomial distribution is given by:

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

Here, $\binom{n}{x}$ represents the number of ways to choose x successes from n trials, p^x is the probability of having x successes, and q^{n-x} is the probability of having $n - x$ failures.

In our economic example, let's say the salesperson has a 20% success rate $(p = 0.2)$ in closing deals. If the salesperson conducts 10 independent trials, the probability of closing exactly 2 deals $(x = 2)$ can be calculated using the binomial distribution PMF. This probability calculation provides insights into the likelihood of achieving a specific number of successful deals in a given number of trials, which is valuable information for sales forecasting and performance evaluation.

Suppose that we have an experiment such as tossing a coin or die repeatedly or choosing a marble from an urn repeatedly. Each toss or selection is called a *trial*. In any single trial there will be a probability associated with a particular event such as head on the coin, 4 on the die, or selection of a red marble. In some cases, this probability will not change from one trial to the next (as in tossing a coin or die). Such trials are then said to be independent and are often called Bernoulli trials after James Bernoulli who investigated them at the end of the seventeenth century.

Let p be the probability that an event will happen in any single Bernoulli trial (called the probability of success). Then $q = 1 - p$ is the probability that the event will fail to happen in any single trial (called the *probability of failure*). The probability that the event will happen exactly x times in n trials (i.e., successes and $n - x$ failures will occur) is given by the probability function

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x} \qquad (1)$$

where the random variable X denotes the number of successes in n trials and $x = 0,1, \ldots, n$.

The discrete probability function $P(X = x)$ for the number of successes in n trials, where $x = 0,1, \ldots, n$, is commonly referred to as the binomial distribution. This distribution is so named because, for each value of x, it corresponds to the coefficients of the binomial expansion of $(q + p)^n$, where q and p are the probabilities of failure and success, respectively. The binomial expansion is a mathematical expression obtained by raising the binomial $(q + p)$ to the power of n.

The binomial distribution formula is given by:

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

This formula encapsulates the probability of obtaining exactly x successes in n independent and identical trials, with each trial having two possible outcomes: success (with probability p) or failure (with probability $q = 1 - p$).

The binomial distribution is further illustrated by the binomial expansion formula, which expands $(q + p)^n$ into a sum of terms, each representing the probability of a specific number of successes. The coefficients $\binom{n}{x}$ in the expansion, correspond to the number of ways to choose x successes from n trials.

In the special case where $n = 1$, the binomial distribution reduces to the Bernoulli distribution, which represents a single Bernoulli trial. The Bernoulli distribution is characterized by the probability of success (p) and the probability of failure $(q = 1 - p)$ in a single trial, making it a fundamental building block for the broader binomial distribution.

Overall, the binomial distribution is a powerful tool in probability theory and statistics, widely used in various fields, including economics, to model and analyze phenomena involving repeated trials with binary outcomes.

The discrete probability function is often called the binomial distribution since for $x = 0, 1, 2, \dots, n$, it corresponds to successive terms in the binomial expansion.

$$(q + p)^n = q^n + \binom{n}{1} q^{n-1}p + \binom{n}{2} q^{n-2}p^2 + \dots + p^n = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} \qquad (2)$$

The special case of a binomial distribution with $n = 1$ is also called the Bernoulli distribution.

Some Properties of The Binomial Distribution

1. **Mean (μ):** The mean of a binomial distribution is given by $\mu = np$, where n is the number of trials and 4.3.2. Binomial Distribution p is the probability of success in a single trial. This provides the average number of successes expected in n trials.

2. **Variance (σ^2):** The variance of a binomial distribution is calculated using

   $\sigma^2 = npq$, where $q = 1 - p$ is the probability of failure.

3. **Standard Deviation (σ):** The standard deviation is the square root of the variance and is given by $\sigma = \sqrt{npq}$.

4. **Moment Generating Function ($M(t)$):** The moment generating function is a function used to derive moments of a distribution. For a binomial distribution, $M(t) = (q + pe^t)^n$, where t is a parameter.

Illustration 4.3.1

Seventy-five percent of employed women say their income is essential to support their family. Let X be the number in a sample of 200 employed women who will say their income is essential to support their family. What is the mean and standard deviation of X

Solution

X follows a binomial distribution with parameters

$\mathrm{n} = 200$ and

$\mathrm{p} = .75$.

The mean is $\mu = \mathrm{np} = 200 \times .75 = 150$, and

The standard deviation is $\sigma = \sqrt{\mathrm{npq}} = \sqrt{37.5} = 6.12$.

Illustration 4.3.2

A binomial distribution has a mean equal to 8 and a standard deviation equal to 2. Find the values for n and p.

Solution:

We use the formulas for the mean and standard deviation of a binomial distribution:

- Mean: $\mu = np = 8$
- Standard deviation: $\sigma = \sqrt{npq} = 2$

First, square the standard deviation:

$\sigma^2 = npq = 2^2 = 4$

Now, use the two equations:

5. $np = 8$
6. $npq = 4$

Substitute equation (1) into equation (2):

$4 = 8q \Rightarrow q = 48 = 0.5$

Since $p + q = 1$, we find:

$p = 1 - 0.5 = 0.5$

Now substitute $p = 0.5$ into the first equation:

$$n \times 0.5 = 8 \Rightarrow n = \frac{8}{0.5} = 16$$

Final Answer:

$n = 16$

$p = 0.5$

Illustration 4.3.3

If on the average rainfall on 10 days in every 30 days, obtain the probability that rain will fall on at least 3 days of a given week.

Solution

The probability density function foe a binomial distribution is

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

Given $p = \frac{10}{30} = \frac{1}{3}, \quad n = 7, \quad q = 1 - \frac{1}{3} = \frac{2}{3}$

$$P[X \geq 3] = 1 - P[X < 3]$$

$$= 1 - P[X = 0,1,2]$$

$$= 1 - [P[X = 0] + P[X = 1] + P[X = 2]]$$

$$P(X = 0) = \binom{7}{0}\left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^{7-0} = \left(\frac{2}{3}\right)^7 = 0.0585$$

$$P(X = 1) = \binom{7}{1}\left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^{7-1} = \binom{7}{1}\left(\frac{2}{3}\right)^6 \left(\frac{1}{3}\right) = 0.2048$$

$$P(X = 1) = \binom{7}{2}\left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^{7-2} = \binom{7}{2}\left(\frac{2}{3}\right)^5 \left(\frac{1}{3}\right)^2 = 0.3073$$

$$P[X \geq 3] = 1 - \{0.0585 + 0.2048 + 0.3073\}$$

$$= 1 - 0.5706 = 0.4293$$

Illustration 4.3.4

Ten coins are thrown simultaneously. Find the probability of getting at least seven heads?

Solution

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

q = Probability of not getting a head $= \frac{1}{2}$

The probability of getting x heads in a random throw of 10 coins is

$$P(x) = \binom{10}{x} p^x q^{10-x} = \binom{10}{x}\left(\frac{1}{2}\right)^{10}, \quad x = 0,1,2 \ldots .10$$

probability of getting at least seven heads = $P[X \geq 7]$

$$= P[X = 7] + P[X = 8] + P[X = 9] + P[X = 10]$$

$$= \binom{10}{7}\left(\frac{1}{2}\right)^{10} + \binom{10}{8}\left(\frac{1}{2}\right)^{10} + \binom{10}{9}\left(\frac{1}{2}\right)^{10} + \binom{10}{10}\left(\frac{1}{2}\right)^{10}$$

$$= \left(\frac{1}{2}\right)^{10}\left[\binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right]$$

$$= \left(\frac{1}{2}\right)^{10}[120 + 45 + 10 + 1] = \frac{176}{1024}$$

Illustration 4.3.5

A die is tossed 3 times. A success is getting 1 or 6 on a toss. Find the mean and variance of the number of successes.

Solution

Given $n = 3,\quad p = p(getting\ 1\ or\ 6) = \frac{1}{6} = \frac{1}{6} = \frac{1}{3}$

Mean = $np = 3 \times \frac{1}{3} = 1$

Variance = $npq = 3 \times \frac{1}{3} \times \frac{2}{3} = \frac{2}{3}$

Illustration 4.3.6

Suppose that a Central University has to form a committee of 5 members from a list of 20 candidates, out of whom 12 are teachers and 8 are students. If the members of the committee are selected at random, what is the probability that the majority of the committee members are students?

Solution

p = Probability of selecting a student member = $\frac{8}{20} = \frac{2}{5}$

q = Probability of selecting a teacher member = $\frac{12}{20} = \frac{3}{5}$

Let X denote the number of students selected in the committee. Hence, by binomial probability distribution,

$$P(x) = \binom{5}{x} p^x q^{5-x} = \binom{5}{x}\left(\frac{2}{5}\right)^x \left(\frac{3}{5}\right)^{5-x}, \quad x = 0,1,2,3,4,5$$

The required probability is given by :

probability of getting at least seven heads = $P[X \geq 3]$

$$= P[X = 3] + P[X = 4] + P[X = 5]$$

$$= \binom{5}{3}\left(\frac{2}{5}\right)^{3}\left(\frac{3}{5}\right)^{5-3} + \binom{5}{4}\left(\frac{2}{5}\right)^{4}\left(\frac{3}{5}\right)^{5-4} + \binom{5}{5}\left(\frac{2}{5}\right)^{5}\left(\frac{3}{5}\right)^{5-5}$$

$$= \left(\frac{2}{5}\right)^{3}\left[10 \times \left(\frac{3}{5}\right)^{2} + 5 \times \left(\frac{2}{5}\right)\left(\frac{3}{5}\right) + \left(\frac{2}{5}\right)^{2}\right]$$

$$= \left(\frac{2}{5}\right)^{3}\left[10 \times \frac{9}{25} + \left(\frac{6}{5}\right) + \left(\frac{4}{25}\right)\right]$$

$$= \left(\frac{2}{5}\right)^{3}\left[\frac{90 + 30 + 4}{25}\right]$$

$$= \left(\frac{2}{5}\right)^{3}\left[\frac{124}{25}\right]$$

$$= \frac{8}{125} \times \frac{124}{25}$$

$$= \frac{992}{3125} = 0.3174$$

4.3.3 Poisson distribution

The Poisson distribution was derived in 1837 by the French mathematician Simeon D. Poisson. Poisson distribution may be obtained as a limiting case of the Binomial probability distribution under the following conditions:

1. n, the number of trials is indefinitely large, i.e., $n \rightarrow \infty$.

2. p, the constant probability of success for each trial is indefinitely small

i.e., $p \rightarrow$

3. The product np = λ remains **finite**, where λ is a fixed constant (mean number of occurrences).

Under the above three conditions, the Binomial probability function tends to the probability function of the Poisson distribution given by

$$p(x) = P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad , x = 0, 1, 2, 3, \ldots$$

where X is the number of successes (occurrences of the event), $\lambda = np$.

Some Properties of The Poission Distribution

1. **Mean (λ):** The mean of a poission distribution is given by $\lambda = np$, where n is the number of trials and p is the probability of success in a single trial.

This provides the average number of successes expected in n trials.

2. **Variance (λ):** The variance of a binomial distribution is calculated using $\sigma^2 = \lambda$,

3. **Standard Deviation (σ):** The standard deviation is the square root of the variance and is given by $\sigma = \sqrt{\lambda}$.

Illustration 4.3.7

If 5% of the electric bulbs manufactured by a company are defective, use Poisson distribution to find the probability that in a sample of 100 bulbs

- none is defective,
- 5 bulbs will be defective. (Given : $e^{-5} = 0 \cdot 007$)

Solution

Given $n = 100$,

$p =$ probability of defective bulbs $= 5\% = \frac{5}{100}$

$\lambda = np = 100 \times \frac{5}{100} = 5$

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, 3, \ldots$$

i. $P(X = 0) = \frac{e^{-5}5^0}{0!} = e^{-5} = 0.007$

ii. $P(X = 5) = \frac{e^{-5}5^5}{5!} = e^{-5} = 0.1823$

Illustration 4.3.7

A manufacturer of cotter pins knows that 5% of his product is defective. If he sells cotter pins in boxes of 100 and guarantees that not more than 10 pins will be defective, what is the approximate probability that a box will fail to meet the guaranteed quality?

Solution

We are given-$n = 100$.

p - Probability of a defective pin $= 5\% = \frac{5}{100} = 0.05$

λ = Mean' number of defective pins in a box of 100

$= np = 100\, x\, 0 \cdot 05$

= 5

Since 'p' is small, we .may use Poisson distribution.

Probability of x detective pins in a box of 100 is

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-5}\,5^x}{x!}, x = 0,1,2,3,\ldots$$

$$P[X > 10] = 1 - P[X \leq 10\,)$$

$$= 1 - \{P[X = 0] + P[X = 1] + \cdots + P[X = 10]$$

$$= 1 - \{\frac{e^{-5}\,5^0}{0!} + \frac{e^{-5}\,5^1}{1!} + \cdots + \frac{e^{-5}\,5^{10}}{10!}\}$$

$$= 1 - e^{-5}\sum_{x=0}^{10}\frac{5^x}{x!}$$

Illustration 4.3.8

The number of accidents in a year to taxi drivers in a city follows a poission distribution with mean 3. Out of 2000 taxi drivers , find the number of drivers with more than 3 accidents in a year?

Solution

We are given-$n = 2000, \lambda = 3$

$$P[X > 3] = 1 - P[X \leq 3\,)$$

$$= 1 - \{P[X = 0] + P[X = 1] + P[X = 2] + P[X = 3]$$

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

$$= 1 - e^{-3}\{1 + 3 + \frac{9}{2} + \frac{27}{6}\}$$

$$= 1 - e^{-3}\left\{1 + 3 + \frac{9}{2} + \frac{27}{6}\right\} = 0.3528$$

Illustration 4.3.9

Suppose that on an average one house in 1000 in a certain district has a fire during a year. If there are 2000 houses in that district what is the probability that exactly 5 houses will have a fire during the year?

Solution

Given-$n = 2000,\ p = \frac{1}{1000}$

$$\lambda = \text{np} = 2000 \times \frac{1}{1000} = 2$$

$$P[X = 5] == \frac{e^{-2}\,2^5}{5!} = \frac{4}{15\,e^2}$$

Illustration 4.3.10

A car hire firm has 2 cars which it hires out day by day. The number of demands for a car on each day is distributed as Piosson distribution with mean 1.5. Calculate the proportion of days on which 1) there is no demand 2) Some demand is refused.

Solution

Given $\lambda = 1.5$

proportion of days with no demand $= P(0) = \frac{e^{-1.5} \times 1.5^0}{0!} = 0.2231$

proportion of days on which some demand is refused = $P[X > 2]$

$$= 1 - \{P[X = 0] + P[X = 1] + P[X = 2]\}$$

$$= 1 - \{\frac{e^{-1.5} \times 1.5^0}{0!} + \frac{e^{-1.5} \times 1.5^1}{1!} + \frac{e^{-1.5} \times 1.5^2}{2!}\}$$

$$= 1 - 0.8087 = 0.1913$$

4.3.4 Uniform Distribution

Consider the random variable X representing the flight time of an aeroplane travelling from Trivandrum to New Delhi. Suppose the flight time can be any value in the interval from 120 minutes to 140 minutes. Because the random variable X can assume any value in that interval, X is a continuous rather than a discrete random variable. Let us assume that sufficient actual flight data are available to conclude that the probability of a flight time within anyone-minute interval is the same as the probability of a flight time within any other one-minute interval contained in the larger interval from120 to140minutes. With everyone-minute interval being equally likely, the random variable X is said to have a uniform probability distribution.

If x is any number lying in the range that the random variable X can take then the probability density function, which defines the uniform distribution for the flight-time random variable, is:

$$f(x) = \frac{1}{20} \qquad for\ 120 \leq x \leq 140$$

$$= 0 \qquad \text{elsewhere}$$

The probability function for uniform distribution is

$$f(x) = \frac{1}{b-a} \qquad for\ a \leq x \leq b$$

$$= 0 \qquad \text{elsewhere}$$

Properties of Uniform Distribution

1. Mean: The mean of a uniform distribution is $\frac{a+b}{2}$, where x is defined for $a \leq x \leq b$

2. Variance: The variance of a uniform distribution is $\frac{(a-b)^2}{12}$.

Illustration 4.3.11

A bus arrives every 15 minutes at a bus stop. Assuming that the waiting time X for the bus is Uniformly distributed, find the probability that a person has to wait for the bus 1) more than 10 minutes 2) between 5 and 10 minutes.

Solution

Given $f(x) = \frac{1}{15} \quad 0 < x < 15$

$= 0, \quad otherwise$

$$P[X > 10] = \int_{10}^{15} \frac{1}{15} \; dx = \frac{1}{15} \int_{10}^{15} \; dx = \frac{1}{15}(15 - 10) = \frac{5}{15} = \frac{1}{3}$$

$$P[5 < X < 10] = \int_{5}^{10} \frac{1}{15} \; dx = \frac{1}{15} \int_{5}^{10} \; dx = \frac{1}{15}(10 - 5) = \frac{5}{15} = \frac{1}{3}$$

Illustration 4.3.12

Subway train on a certain line run every half hour between mid-night and six in the morning. What is the probability that a man entering the station at a random time during this period will have to wait at least 20 minutes?

Solution

Let the random variable X denote the waiting time (in minutes) for the next train. Under the assumption that a man arrives at the station at random, X is distributed uniformly on (0, 30), with probability function

$$f(x) = \frac{1}{30} \quad 0 < x < 30$$

The probability that he has to wait at 'east 20 minutes is

$$[X > 20] = \int_{20}^{30} \frac{1}{30} \; dx = \frac{1}{30} \int_{20}^{30} \; dx = \frac{1}{30}(30 - 20) = \frac{10}{30} = \frac{1}{3}$$

Illustration 4.3.13

A shuttle train at a busy airport completes a circuit between 2 terminals every 5 minutes. What is the probability that a passenger will wait more than 3 minutes for a shuttle train

Solution

Let the random variable X denote the waiting time (in minutes) for the shuttle train.

X is distributed uniformly on (0, 5), with probability function

$$f(x) = \frac{1}{5}$$

The probability that he has to wait more than 3 minutes is

$$[X > 3] = \int_3^5 \frac{1}{5}\, dx = \frac{1}{5} \int_3^5 dx = \frac{1}{5}(5-3) = \frac{2}{5}$$

Summarised Overview

Discrete variables are random variables that take distinct, separate values, often associated with countable or finite outcomes in probability distributions. One important discrete distribution is the binomial distribution, which models the number of successes in a fixed number of independent Bernoulli trials. This distribution is characterized by two key parameters: the number of trials and the probability of success in each trial. Another key distribution is the Poisson distribution, which is a special case of the binomial distribution. It is used to model rare events, especially when the number of trials is very large and the probability of success in each trial is very small. Lastly, the uniform distribution describes a situation where all outcomes in a given range are equally likely. The probability density function (PDF) of this distribution is constant across the range that the random variable can take. These distributions are fundamental in understanding and modelling various types of random processes.

Assignments

1. Thirty percent of the trees in a forest are infested with a parasite. Fifty trees are randomly selected from this forest and $\mathtt{X}$ is defined to equal the number of trees in the 50 sampled that are infested with the parasite. The infestation is uniformly spread throughout the forest. Identify the values for $\mathtt{n}$, $\mathtt{p}$, and $\mathtt{q}$. Suppose we define $\mathtt{Y}$ to be the number of trees in the 50 sampled that are not infested with the parasite. Then $\mathtt{Y}$ is a binomial random variable.

   a. What are the values of $\mathtt{n}$, $\mathtt{p}$, and $\mathtt{q}$ for $\mathtt{Y}$?

   b. The event $\mathtt{X} = 20$ is equivalent to the event that $\mathtt{Y} = \mathtt{a}$. Find the value for $\mathtt{a}$.

2. The mean of a binomial distribution is 4 and its standard deviation is $\sqrt{3}$. What are the values of n, p and q with usual notations? $n = 16, p = \frac{1}{4},\ q = \frac{3}{4}$

3. In a Binomial distribution with 6 independent trials, the probabilities of 3 and 4 successes are found to be 0·2457 and 0·0819, respectively. Find the parameter 'p' of the Binomial distribution.

4. A manufacturer of blades knows that 5% of his product is defective. If he sells blades in boxes of 100 and guarantees that not more than 10 blades will be defective, what is the probability (approximately) that a box will fail to meet the guaranteed quality?

5. In a certain factory turning out optical lenses, there is a small chance of 1/500 for any one lens to be defective. The lenses are supplied in packets of 10. Use the Poisson distribution to calculate the approximate number of packets containing no defective, one defective, two defective, three defective lenses, respectively, in a consignment of 20,000 packets. You are given that

   $e^{-0.02} = 0 \cdot 9802$.

6. A bus arrives every 30 minutes at a bus stop. Assuming that the waiting time X for the bus is uniformly distributed, find the probability that a person has to wait for the bus 1) more than 10 minutes 2) between 5 and 10 minutes.

References

1. Gujarathi, D. Sangeetha, N. (2007). *Basic Econometrics* (4th ed.) New Delhi: McGraw-Hill

2. Koutsoyianis, A. (1977*). Theory of Econometrics (2nd ed.*). London. The Macmillan Press Ltd

Suggested Readings

1. Anderson, D., D. Sweeney and T. Williams (2013): *"Statistics for Business and Economics*", Cengage Learning: New Delhi.

2. Goon, A.M., Gupta, and Das Gupta B (2002): *Fundamentals of Statistics* (Vol I), World Press.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

4

UNIT

Continuous Probability Distributions: Foundations and Applications

Learning Outcomes

After completing this unit, learner will be able to:

- interpret continuous probability distributions
- differentiate between normal, lognormal, and exponential distributions

Background

In today's data-driven world, understanding uncertainty and variability is crucial for making informed decisions, and probability distributions are essential mathematical tools for modeling and analyzing random phenomena. With the increasing availability of large datasets and computational power, probability distributions have become a fundamental component of modern data science and analytics, applied in various fields, including business, engineering, economics, finance, and healthcare. By studying probability distributions, individuals can gain valuable skills in data analysis, statistical modeling, and decision-making under uncertainty, driving success in their careers and contributing to advancements in their respective fields.

Keywords

Continuous Probability Function, Normal Distribution, Lognormal Distributions, Exponential Distribution

4.4.1 Continuous Probability Function

Random variables can take on different forms based on the nature of the outcomes they represent. A discrete random variable is one that can assume a finite or countably infinite set of distinct values. For instance, the number of students in a classroom or the count of defective items in a production batch are examples of discrete random variables. On the other hand, a no discrete random variable takes on a noncountably infinite number of values and often corresponds to continuous phenomena, such as the height of individuals or the temperature in a given location.

A function $f(x)$ that satisfies the above requirements is called a *probability function* or *probability distribution* for a continuous random variable, but it is more often called a *probability density function* or *simply density function*. Any function $f(x)$ satisfying Properties 1 and 2 above will automatically be a density function.

A random variable X is said to be absolutely continuous, or simply continuous, if the Probability density function is

1. $f_i(x_i) \geq 0$
2. $\int_{-\infty}^{\infty} f_i(x_i)\, dx_i = 1$

4.4.2 Normal Distribution

One of the most important examples of a continuous probability distribution is the *normal distribution*, sometimes called the *Gaussian distribution*. The density function for this distribution is given by

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

where μ and σ are the mean and standard deviation, respectively. The corresponding distribution function is given by

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(v-\mu)^2/2\sigma^2} dv$$

If X has the distribution function, we say that the random variable X is *normally distributed* with mean μ and variance σ^2.

If we let Z be the standardized variable corresponding to X, i.e., if we let

$$z = \frac{x-\mu}{\sigma} \approx (0,1),$$

then the mean or expected value of Z is 0 and the variance is 1. In such cases the density function for Z can be formally placing $\mu = 0$ and $\sigma = 1$, yielding

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

This is often referred to as the *standard normal density function*.

A graph of the density function, sometimes called the *standard normal curve*. In this graph we have indicated the areas within 1, 2, and 3 standard deviations of the mean (i.e., between $z = -1$ and $+1$, $z = -2$ and $+2$, $z = -3$ and +3) as equal, respectively, to 68.27%, 95.45% and 99.73% of the total area, which is one. This means that,

$$P(-1 \leq Z \leq 1) = 0.6827$$
$$P(-2 \leq Z \leq 2) = 0.9545$$
$$P(-3 \leq Z \leq 3) = 0.9973$$

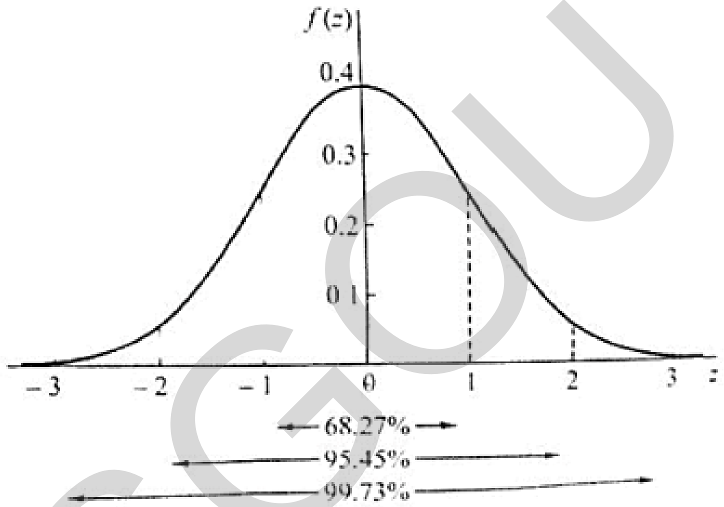
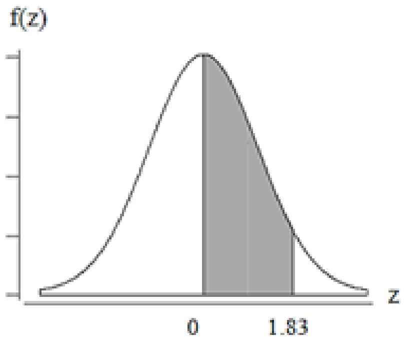
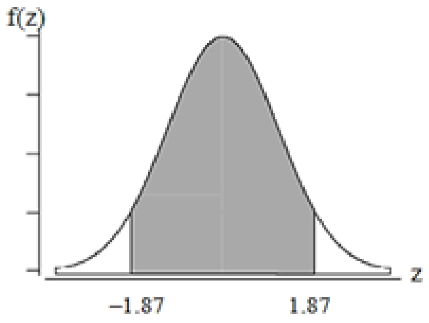


Illustration 4.4.1

Express the areas shown in the following two standard normal curves as a probability statement and find the area of each one.





Solution

The area under the curve on the left is represented as $P(0 < Z < 1.83)$ and from the standard normal distribution table is equal to 0.4664. The area under the curve on the right is represented as $P(-1.87 < Z < 1.87)$ and from the standard normal distribution table is $2 \times 0.4693 = 0.9386$.

Illustration 4.4.1

The distribution of complaints per week per 100,000 passengers for all airlines in a country is normally distributed with $\mu = 4.5$ and $\sigma = 0.8$. Find the standardized values for the following observed values of the number of complaints per week per 100,000 passengers: (a) 6.3; (b) 2.5; (c) 4.5; (d) 8.0.

Solution

(a) The standardized value for 6.3 is found by $z = \frac{x-\mu}{\sigma} = \frac{6.3-4.5}{.8} = 2.25$

(b) The standardized value for 2.5 is found by $z = \frac{x-\mu}{\sigma} = \frac{2.5-4.5}{.8} = -2.50$

(c) The standardized value for 4.5 is found by $z = \frac{x-\mu}{\sigma} = \frac{4.5-4.5}{.8} = 0.00$

(d) The standardized value for 8.0 is found by $z = \frac{x-\mu}{\sigma} = \frac{8.0-4.5}{.8} = 4.38$

Illustration 4.4.2

The net worth of senior citizens is normally distributed with mean equal to $\$225{,}000$ and standard deviation equal to $\$35{,}000$. What percent of senior citizens have a net worth less than $\$300{,}000$?

Solution

Ans. Let X represent the net worth of senior citizens in thousands of dollars. The percent of senior citizens with a net worth less than $\$300{,}000$ is found by multiplying $P(X < 300)$ times 100. The probability $P(X < 300)$ is shown in figure below. The event $X < 300$ is equivalent to the event $Z < \frac{300-225}{35} = 2.14$. The probability that $Z < 2.14$ is represented as the shaded area in Fig. 4.4. The probability that Z is less than 2.14 is found by adding $P(0 < Z < 2.14)$ to .5, which equals. $5 + .4838 = .9838$. We can conclude that 98.38% of the senior citizens have net worth less than$\$300{,}000$.

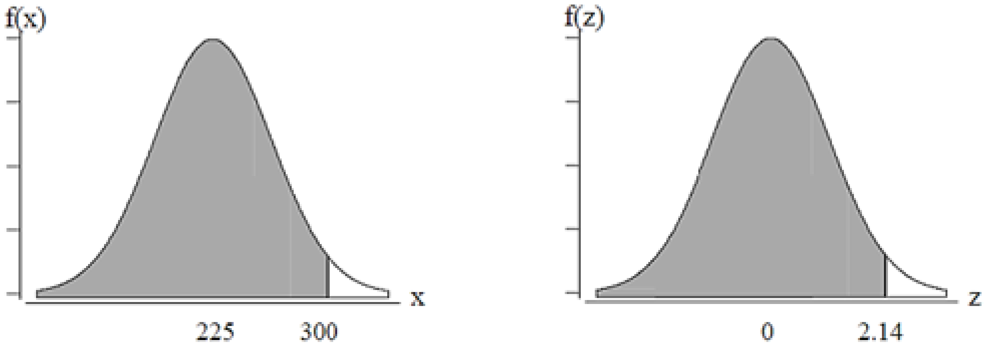


Illustration 4.4.3

The average test marks in a particular class is 79 and standard deviation is 5. If the marks are normally distributed, how many students in a class of 200 did not receive marks between 75 and 82?

Solution

Given $\mu = 79, \ \sigma = 5, \ n = 200$

$$z = \frac{x - \mu}{\sigma} = \frac{75 - 79}{5} = -0.8$$

$$z = \frac{x - \mu}{\sigma} = \frac{82 - 79}{5} = 0.6$$

$$P[75 < X < 82] = P[-0.8 < Z < 0.6]$$

$$= 0.2881 + 0.2257 = 0.5138$$

4.4.3 Lognormal Distributions

A log-normal distribution is a probability distribution of a random variable whose logarithm is normally distributed. In other words, if a random variable *X* follows a log-normal distribution, then *ln X* (the natural logarithm of *X* follows a normal distribution.

The Probability density function of a log-normal random variable *X* is:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(ln(x)-\mu)^2}{2\sigma^2}} \quad x > 0$$

If X follows lognormal distribution lognormal $N(\mu,\sigma^2)$ then $ln(X)$ follows normal distribution $N(\mu,\sigma^2)$

4.4.4 Exponential Distribution

A continuous random variable X as with non-negative values is said to have an exponential distribution with parameter $\lambda > 0$, if its probability density function is given by

$$f(x) = \lambda\, e^{-\lambda x} \quad if\ \lambda > 0$$

$$= 0 \quad if\ \lambda \leq 0$$

Some Properties of the Exponential Distribution

1. Mean

The mean of an exponential distribution is given by $\frac{1}{\lambda}$ where λ is the parameter.

2. Variance

The variance of an exponential distribution is $\frac{1}{\lambda^2}$.

Illustration 4.4.4

The amount of time that a watch will run without having to be reset is a random variable having an exponential distribution with mean 120 days. Find the probability that such a watch will

i. have to reset in less than 24 hours

ii. not have to reset in at least 180 days?

Solution

Given mean $\frac{1}{\lambda} = 120$

The probability density function is given by

$$f(x) = \lambda\, \mathrm{e}^{-\lambda x} \quad if\ \lambda > 0$$

$$= \frac{1}{120}\, \mathrm{e}^{-\frac{1}{120}x}$$

i. P [watch have to be reset in less than 24 hours]

= P[watch will run for less than 24 days]

$$= P[X < 24]$$

$$= \int_0^{24} f(x)\, dx$$

$$= \int_0^{24} \frac{1}{120}\, \mathrm{e}^{-\frac{1}{120}x} dx$$

$$= \frac{1}{120} \int_0^{24} \mathrm{e}^{-\frac{1}{120}x} dx$$

$$= \frac{1}{120} \left(\frac{\mathrm{e}^{-\frac{1}{120}x}}{-\frac{1}{120}} \right)_0^{24}$$

$$= \left(\mathrm{e}^{-\frac{24}{120}} - \mathrm{e}^{0} \right) = e^{-\frac{1}{5}} - 1$$

Illustration 4.4.5

The milage which a car owner gets with a certain kind of tyre is a random variable having exponential distribution with mean 40,000 kms. Find the probability that one of these tyres will last i) at least 30000 kms 2) at most 35,000 kms.

Solution

Given mean $\frac{1}{\lambda} = 40{,}000$

The probability density function is given by

$f(x) = \lambda\, e^{-\lambda x}\ \ if\ \lambda > 0$

$$= \frac{1}{40000}\, e^{-\frac{1}{40000}x}$$

i. P [one of these tyres will last at least 30000 kms]

$$= P[X > 30000]$$

$$= \int_{30000}^{\infty} \frac{1}{40000}\, e^{-\frac{1}{40000}x}\, dx$$

$$= \frac{1}{40000}\int_{30000}^{\infty} e^{-\frac{1}{40000}x}\, dx$$

$$= \frac{1}{40000}\left(\frac{e^{-\frac{1}{40000}x}}{-\frac{1}{40000}}\right)_{30000}^{\infty}$$

$$= e^{-\frac{30000}{40000}}$$

$$= e^{-\frac{3}{4}}$$

Illustration 4.4.6

The time in hours required to repair a machine is exponentially distributed with $\lambda = \frac{1}{20}$.

What is the probability that the required time

i. Exceeds 30 hrs.

ii. In between 16 hrs. and 24 hrs.

iii. At most 10 hrs

Solution

$f(x) = \lambda\, e^{-\lambda x}\ \ if\ \lambda > 0$

$$= \frac{1}{20}\, e^{-\frac{1}{20}x} \quad x > 0$$

i. $P[X] > 30 = \int_{30}^{\infty} \frac{1}{20}\, e^{-\frac{1}{20}x}\, dx$

$$= \frac{1}{20}\left(\frac{e^{-\frac{1}{20}x}}{-\frac{1}{20}}\right)_{30}^{\infty}$$

$$= -e^{-\frac{30}{20}} = e^{-\frac{3}{2}}$$

ii. $P[16 < X < 24]$

$$= \int_{16}^{24} f(x)\,dx$$

$$= \int_{16}^{24} \frac{1}{20}\,e^{-\frac{1}{20}x}dx$$

$$= \frac{1}{20}\left(\frac{e^{-\frac{1}{20}x}}{-\frac{1}{20}}\right)_{16}^{24}$$

$$= -\left(e^{-\left(\frac{24}{20}\right)} - e^{-\left(\frac{16}{20}\right)}\right)$$

$$= \left(e^{-\left(\frac{6}{5}\right)} - e^{-\left(\frac{4}{5}\right)}\right)$$

ii. $P[X \le 10]$

$$= \int_{0}^{10} \frac{1}{20}\,e^{-\frac{1}{20}x}dx$$

$$= \frac{1}{20}\left(\frac{e^{-\frac{1}{20}x}}{-\frac{1}{20}}\right)_{0}^{10}$$

$$= -\left(e^{-\left(\frac{10}{20}\right)} - 1\right)$$

$$= 1 - e^{-\frac{1}{2}}$$

Summarised Overview

Continuous variables are random variables that can take any value within a range, often associated with measurements and infinite possible outcomes, characterized by a continuous probability distribution. The normal distribution, also known as the Gaussian distribution or bell curve, is a continuous probability distribution characterized by a symmetric, bell-shaped curve, where the majority of observations cluster around the mean. The probability distribution of a random variable whose logarithm is normally distributed is log-normal distribution. The exponential distribution is a continuous probability distribution often used to model the time between events in a Poisson process. It describes situations where events occur independently and at a constant average rate.

Assignments

1. The hospital cost for individuals involved in accidents who do not wear seat belts is normally distributed with mean Rs. 7500 and standard deviation Rs. 1200.

   (a) Find the cost for an individual whose standardized value is 2.5.

   (b) Find the cost for an individual whose bill is 3 standard deviations below the average.

2. The average TV viewing time per week for children ages 2 to 11 is 22.5 hours and the standard deviation is 5.5 hours. Assuming the viewing times are normally distributed, find the following.

   a. What percent of the children have viewing times less than 10 hours per week?

   b. What percent of the children have viewing times between 15 and 25 hours per week?

   c. What percent of the children have viewing times greater than 40 hours per week?

3. In a certain examination 15% of the candidates passed with distinction while 25% of them failed. It is known that a candidate fails if he obtains less than 40 marks (out of 100) while he must obtain at least 75 marks in order to pass with distinction. Determine mean and standard deviation of the distribution of marks assuming this to be normal.

4. If the cauliflowers on a truck are classified as A, B and C according to a size-weight index as: under 75, between 75 and 80, and above 80; find

approximately (assuming a normal distribution) the mean and standard deviation of a lot in which A are 58%, B are 38% and C are 4%.

5. The time in hours required to repair a machine is exponentially distributed with $\lambda = \frac{1}{30}$. What is the probability that the required time

   Exceeds 20 hrs.

   In between 18 hrs. and 22 hrs.

   At most 12 hrs.

References

1. Gujarathi , D.&Sangeetha, N. (2007). *Basic Econometrics* (4thed) New Delhi: McGraw Hill

2. Koutsoyianis, A. (1977). *Theory of Econometrics* (2nded). London .The Macmillian Press Ltd

Suggested Readings

1. Anderson, D., D.Sweeney and T.Williams (2013): "*Statistics for Business and Economics*", Cengage Learning : New Delhi.

2. Goon, A.M. , Gupta and Das Gupta B (2002): *Fundamentals of Statistics* (Vol I), World Press.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

സർവ്വകലാശാലാഗീതം

വിദ്യയാൽ സ്വതന്ത്രരാകണം
വിശ്വപൗരരായി മാറണം
ഗ്രഹപ്രസാദമായ് വിളങ്ങണം
ഗുരുപ്രകാശമേ നയിക്കണേ

കൂരിരുട്ടിൽ നിന്നു ഞങ്ങളെ
സൂര്യവീഥിയിൽ തെളിക്കണം
സ്നേഹദീപ്തിയായ് വിളങ്ങണം
നീതിവൈജയന്തി പാറണം

ശാസ്ത്രവ്യാപ്തിയെന്നുമേകണം
ജാതിഭേദമാകെ മാറണം
ബോധരശ്മിയിൽ തിളങ്ങുവാൻ
ജ്ഞാനകേന്ദ്രമേ ജ്വലിക്കണേ

കുരീപ്പുഴ ശ്രീകുമാർ

SREENARAYANAGURU OPEN UNIVERSITY

Regional Centres

Kozhikode

Govt. Arts and Science College
Meenchantha, Kozhikode,
Kerala, Pin: 673002
Ph: 04952920228
email: rckdirector@sgou.ac.in

Thalassery

Govt. Brennen College
Dharmadam, Thalassery,
Kannur, Pin: 670106
Ph: 04902990494
email: rctdirector@sgou.ac.in

Tripunithura

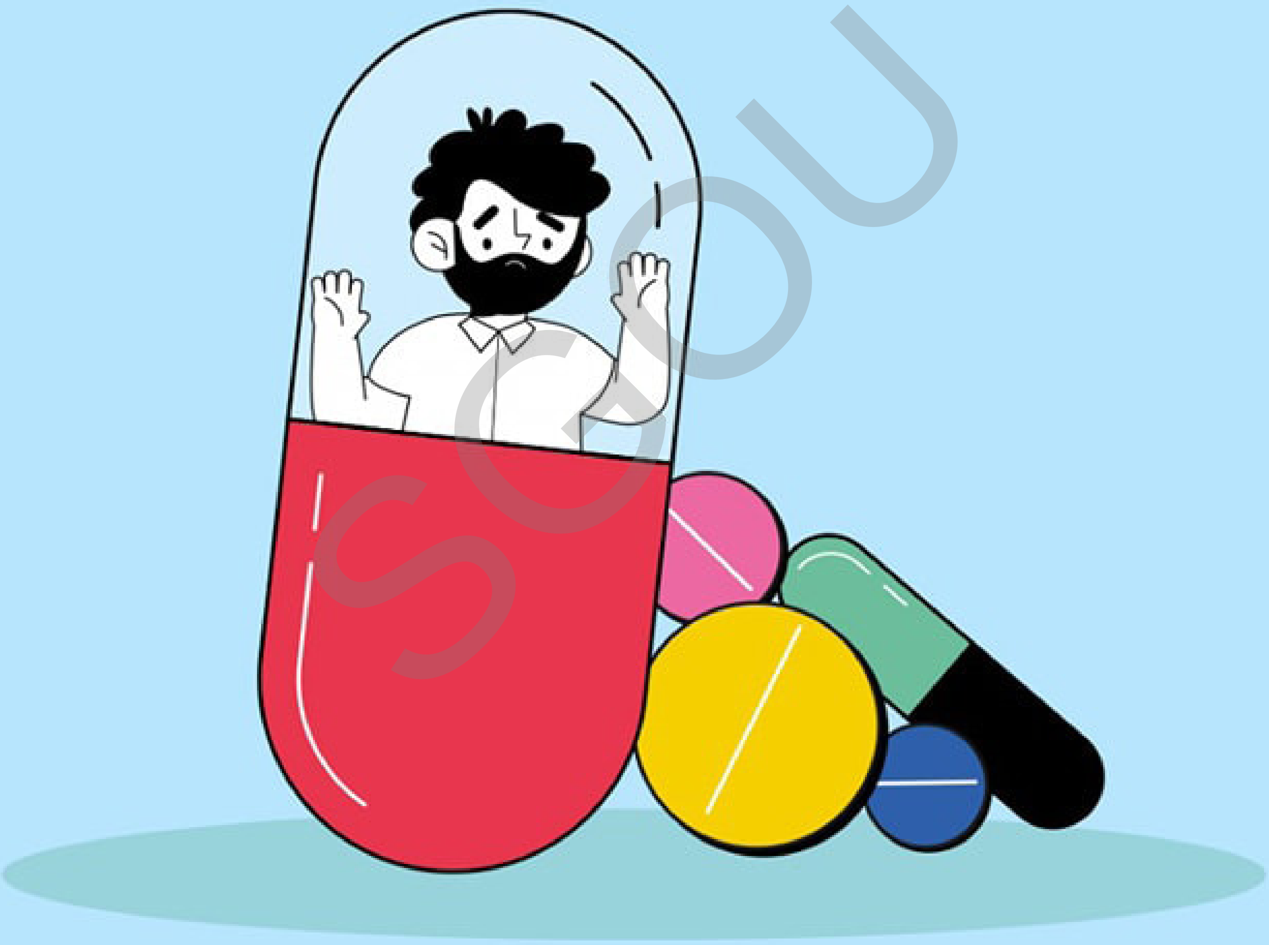
Govt. College
Tripunithura, Ernakulam,
Kerala, Pin: 682301
Ph: 04842927436
email: rcedirector@sgou.ac.in

Pattambi

Sree Neelakanta Govt. Sanskrit College
Pattambi, Palakkad,
Kerala, Pin: 679303
Ph: 04662912009
email: rcpdirector@sgou.ac.in

NO TO DRUGS

തിരിച്ചിറങ്ങാൻ പ്രയാസമാണ്



NO TO DRUGS

ആരോഗ്യ കുടുംബക്ഷേമ വകുപ്പ്, കേരള സർക്കാർ

QUANTITATIVE METHODS FOR ECONOMICS II

COURSE CODE: M23EC10DC

YouTube

Sreenarayanaguru Open University
Kollam, Kerala Pin- 691601, email: info@sgou.ac.in, www.sgou.ac.in Ph: +91 474 2966841

ISBN 978-81-985949-4-5

9 788198 594945