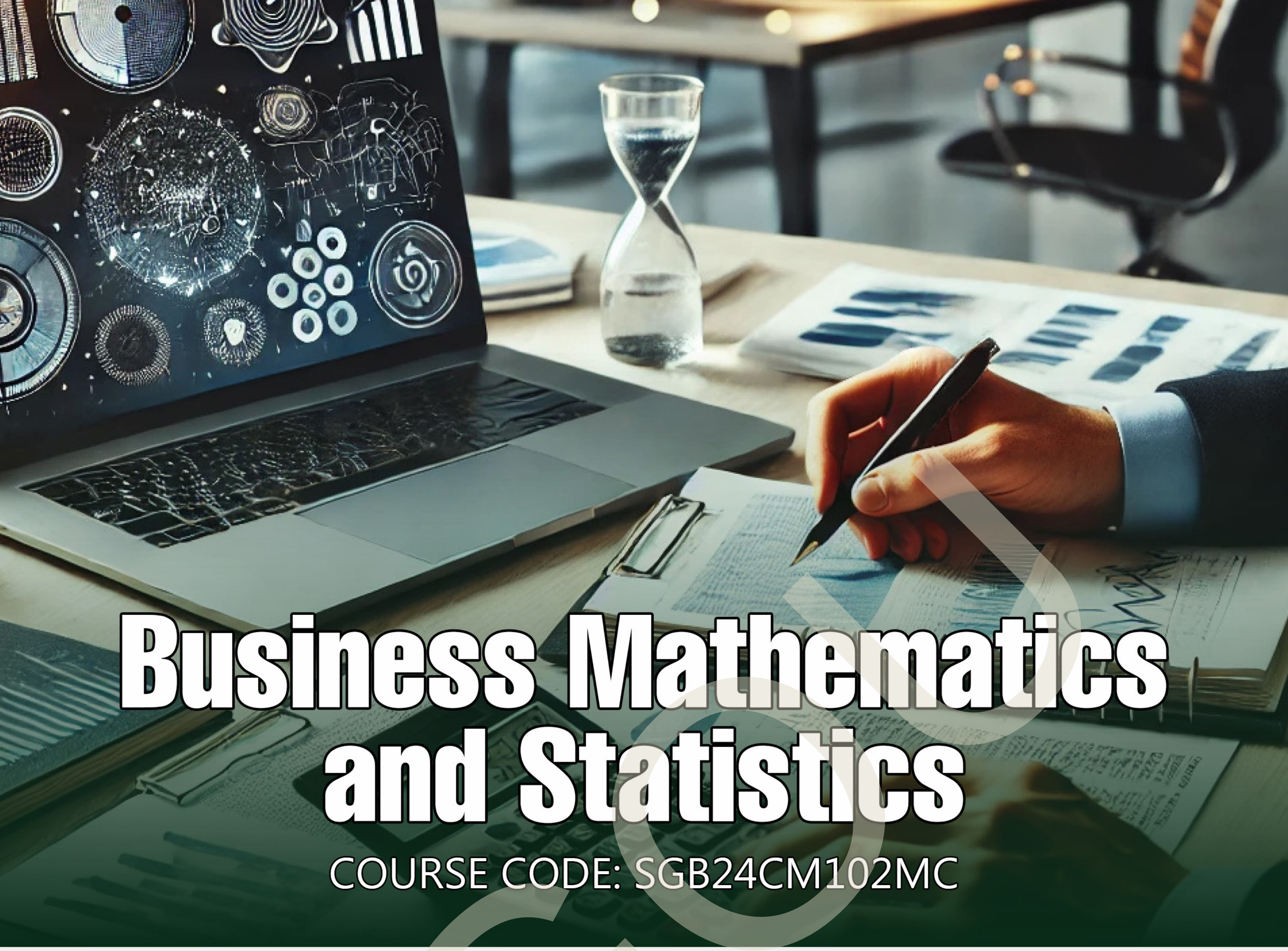


# Business Mathematics and Statistics

COURSE CODE: SGB24CM102MC

Bachelor of Commerce (Honours)
Major Course
Self Learning Material

വിദ്യകൊണ്ട് സ്വതന്ത്രരാവുക
SREENARAYANAGURU
OPEN UNIVERSITY

## SREENARAYANAGURU OPEN UNIVERSITY

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

## 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.

# Business Mathematics and Statistics

Course Code: SGB24CM102MC

Semester - II

## Four Year Undergraduate Programme
## Bachelor of Commerce (Honours)
## Major Course
## Self Learning Material



SREENARAYANAGURU
OPEN UNIVERSITY

**SREENARAYANAGURU OPEN UNIVERSITY**

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

# BUSINESS MATHEMATICS AND STATISTICS

Course Code: SGB24CM102MC
Semester- II
Major Course
Bachelor of Commerce (Honours)

SREENARAYANAGURU OPEN UNIVERSITY




**Development of the Content**

Dr. Vikraman Pillai
Dr. Sanjith S.R.
Dr. Alice Mani
Dr. Midhun V.

**Review and Edit**

Dr. Anitha C.S.
Dr. V. Jayaraju
Dr. Krishnadas K.

**Linguistics**

Dr. Anitha C.S.

**Scrutiny**

Dr. Bino Joy, Dr. Gauri. L., Dr. Midhun V., Dr. Sanjith S.R., Dr. Dhanya K.M., Mahesan K. P., Dr. Athena Prince, Dr. Gopika C.G., Dr. Kavitha S., Dr. Dany Thomas

**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:**
January 2025





ISBN 978-81-984516-0-6

# MESSAGE FROM VICE CHANCELLOR

Dear Learner,

It is with great pleasure that I welcome you to the Four Year BCom Programme offered by Sreenarayanaguru Open University.

Established in September 2020, our university aims to provide high-quality higher education through open and distance learning. Our guiding principle, 'access and quality define equity', shapes our approach to education. We are committed to maintaining the highest standards in our academic offerings.

Our university proudly bears the name of Sreenarayanaguru, a prominent Renaissance thinker of modern India. His philosophy of social reform and educational empowerment serves as a constant reminder of our dedication to excellence in all our academic pursuits.

The Four Year BCom Programme covers all relevant areas aligned with modern business practices and economic principles. We have incorporated the latest trends in commercial studies to ensure a comprehensive and up-to-date curriculum. Moreover, the programme encompasses flexible options for learners to choose from a range of Ability Enhancement Courses, Multi-disciplinary Courses, Value Added Courses, and Skill Enhancement Courses, complemented by discipline-oriented Advanced and Additional Advanced Courses.

Our teaching methodology combines three key elements: Self Learning Material, Classroom Counselling, and Virtual modes. This blended approach aims to provide a rich and engaging learning experience, overcoming the limitations often associated with distance education. We are confident that this programme will enhance your understanding of commercial principles and practices, preparing you for various career paths and further academic pursuits.

Our learner support services are always available to address any concerns you may have during your time with us. We encourage you to reach out with any questions or feedback regarding the programme.

We wish you success in your academic journey with Sreenarayanaguru Open University.

Best regards,

Dr. Jagathy Raj V.P.
Vice Chancellor
Sreenarayanaguru Open University

01-01-2025

# Table of CONTENTS



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

# 1 BLOCK

# Arithmetic: Computation of Interest

# Unit - 1
# Interest Calculation

## Learning Outcomes

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

- ✓ understand the difference between simple and compound interest
- ✓ compute simple Interest and compound interest for various time periods
- ✓ apply interest computation in daily life situations

## Prerequisites

One day, Ravi's friend, Anya, came to him. She needed **₹1,000** to buy seeds for her farm. Ravi agreed to lend her the money but said,

"You'll have to give me a small extra amount in return for using my money. Let's call it **interest**. You can pay me back in a year with an extra **10% of ₹1,000**."

Anya was puzzled. "What does 10% mean?" she asked.

Ravi explained, "It's a small part of the total amount I gave you. **10% means 10 out of every 100.** So, 10% of ₹1,000 is ₹100. You'll return ₹1,000 plus ₹100, which is ₹1,100, after a year."

Anya smiled. "That's fair. I'll return ₹1,100 in one year!"

She worked hard, grew a lot of crops, and sold them. At the end of the year, she gave Ravi ₹1,100 as promised. She thanked Ravi for the help, and he explained further, "If I lend you money for more years, the interest keeps adding up each year."

Anya nodded and shared what she learned with other villagers. Soon, everyone in the village understood how borrowing and interest worked - and they always planned carefully when borrowing money.

## Keywords

Simple interest, Compound interest, Effective yield

## Discussion

### 1.1.1 Calculation of interest

Interest on a deposit is payable to the depositor periodically from the date of deposit or at maturity along with the principal. But on request from the depositor, interest may be paid at monthly, half-yearly or yearly intervals in case of fixed deposits of twelve months and above.

Basically, there are two types of calculation of interest. These two types are explained in the following tables.

Table (1):

| Year | Principal | Interest 10% | Total Interest Accumulated | Amount |
|---|---|---|---|---|
| 1 | 10,000 | 1000 | 1000 | 11000 |
| 2 | 10,000 | 1000 | 2000 | 12000 |
| 3 | 10,000 | 1000 | 3000 | 13000 |

Table (2):

| Year | Principal | Interest 10% | Amount |
|---|---|---|---|
| 1 | 10000 | 1000 | 11000 |
| 2 | 11000 | 1100 | 12100 |
| 3 | 12100 | 1210 | 13310 |

In table 1 there is the same interest in each year and principal is also same at beginning of each year. In table (2) interest calculation is entirely different from the Table (1). For the first-year interest calculation is same in Table (1) and (2). In table (2), second year interest is calculated on first year interest together with principal. That is in table (2) interest is calculated for the amount and interest for just preceding year. Table (1) and Table (2) show simple interest calculation and compound interest calculation on ₹10000 at 10% per annum for 3 years.

The amount borrowed by a businessman for his business needs is called the principal amount (P). Similarly in the case of a depositor, the amount that is deposited in the bank is also called Principal amount. The bank will charge interest (I) at a specified rate(r) on Principal between the loan date and the repayment date or the period of amount deposited to the bank up to its maturity. This period of time is called interest period (t).

## 1.1.2 Simple Interest (I)

Calculation of simple interest is very easy. The interest rate is a percent of the principal and is usually is an annual (yearly) rate. Most loans or deposits are not for exactly in years. Period of loan may be required in months and day also.

The fundamental formula for simple interest is
Simple Interest = Principal × Rate × Time

$$I = p\,r\,t$$

Where $I$-interest, $p$ – Principal amount, $r$ – rate of interest, $t$ – time period.

**Converting interest rate**
To convert an annual rate to a monthly rate, divide the annual rate by 12. To convert a monthly rate to an annual rate, multiply the monthly rate by 12.

Example: (i) 0.60% per month = 0.60×12 =7.2% per year
(ii) 1% per month = 1 × 12 = 12% per year
(iii) 18% per year = $\frac{18}{12}$ = 1.5 % per month
(iv) 30% per year = $\frac{30}{12}$ = 2.5% per month

Remarks: Simple interest formula has 4 variables: Interest, Principal, Rate and Time. Out of these four variables, if any three variables are given, we can always compute the fourth by changing the formula as follows:

$$I = p\,r\,t$$
$$P = \frac{I}{\mathrm{rt}}$$

$$r = \frac{I}{pt}$$

$$t = \frac{I}{pr}$$

**Illustration 1.1.1**

Find simple interest of ₹60000 at a rate of 12% for

i. 3 years
ii. $\frac{1}{2}$ year
iii. 15 months
iv. 9 months
v. 45 days

**Solutions**

Principal p = ₹60000

Rate of interest $r = 12\%$

i. Simple Interest $I = p\,r\,t$

$$I = 60000 \times \frac{12}{100} \times 3$$

= ₹ 21600

ii. $I = p\,r\,t$

$$= 60000 \times \frac{12}{100} \times \frac{1}{2}$$

= ₹ 3600

iii. The time period is given in months. Therefore, before computing the interest, change the time into years by dividing the number of months by 12.

$$I = p\,r\,t$$

$$= 60000 \times \frac{12}{100} \times \frac{15}{12} = ₹\ 9000$$

iv. $I = p\,r\,t$

$$= 60000 \times \frac{12}{100} \times \frac{9}{12} = ₹\ 5400$$

v. The time period is given in days. Therefore, before computing the interest, change the time into years by dividing the number of days by 365

$$I = p\,r\,t$$

$$= 60000 \times \frac{12}{100} \times \frac{45}{365} = ₹\,887.67$$

**Illustration 1.1.2**

Find Simple Interest and amount on ₹10000 at 10% per annum for 3 years?

**Solution**

$$S.I = p\,r\,t$$

$$= 10000 \times \frac{10}{100} \times 3 = ₹3000$$

Amount = Principal + Interest

= 10000 + 3000 = ₹13000

**Illustration 1.1.3**

Find Simple Interest on ₹3000 at 6% for $4\frac{1}{2}$ years.

**Solutions**

P = 3000

$r = \frac{6}{100}$

t = 4.5

$S\,I = prt = 3000 \times \frac{6}{100} \times 4.5 =$ ₹810

**Illustration 1.1.4**

If ₹1,50,000 is invested at 15% simple interest per year or annually, how much interest is earned after 9 months?

**Solution**

Here time is in terms of months but interest is in terms of years. So, time must be expressed in the same units that in the rate.

$p = 150000$

$r = \frac{15}{100}$

$t$ = 9 months = $\frac{9}{12}$ years

$I = p\,r\,t = 150000 \times \frac{15}{100} \times \frac{9}{12} =$ ₹16875

**Illustration 1.1.5**

If Simple Interest on a certain sum at 8% per annum for 4 years is ₹800, what is the principal amount?

**Solution**

$I = p\,r\,t$

$p = \frac{I}{rt} = \frac{800}{\left(\frac{8}{100} \times 4\right)} = \frac{800 \times 100}{(8 \times 4)} =$ ₹ 2500

**Illustration 1.1.6**

At what per cent per annum will ₹4000 earn ₹1000 simple interest in 5 years.

**Solution**

$I = p\,r\,t$

$r = \frac{I}{pt} = \frac{1000}{4000 \times 5} = \frac{1}{20} = 0.05 = 5\%$

**Illustration 1.1.7**

In how many years will ₹10000 amount to ₹13000 at 10% per annum simple interest.

**Solution**

Here simple interest is not directly given. We have,

Simple Interest = Amount – Principal

$I = 13000 - 10000 =$ ₹3000

$I = p\,r\,t$

$t = \frac{I}{pr} = \frac{3000}{10{,}000 \times \left(\frac{10}{100}\right)} = \frac{3000 \times 100}{10{,}000 \times 10} = 3\ years$

**Formula**

Simple Interest differs from year to year. If $R_1$ is the rate of interest for 1st year, $R_2$ is the rate of interest for 2nd year and $R_3$ is the rate of interest for 3rd year ⋯

Then, Simple Interest = $\frac{P(R_1+R_2+R_3)}{100}$

Here t = 1 year

**Illustration 1.1.8**

Find simple interest on ₹2000 for 3 years if the rate of interest for first year is 4%, interest for second year is 6% and that of third year is 7% respectively.

**Solution**

Simple interest = $\frac{P(R_1+R_2+R_3)}{100}$

$= \frac{2000(4+6+7)}{100}$

$= \frac{2000 \times 17}{100}$

= ₹340

## 1.1.3 Compound Interest

After grasping the fundamentals of simple interest, it's time to explore a more dynamic concept: compound interest. Unlike simple interest, which calculates interest solely on the initial principal, compound interest takes into account the interest that accumulates on previously earned interest, making it a powerful tool for growing investments and savings over time.

We have acquired the skill of calculating compound interest over extended periods, such as 2 or 3 years, by first determining the amount at the end of the initial year and subsequently at the conclusion of each subsequent year, as the situation necessitates.

Amount = Principal $(1 + Rate)^{Time}$

i.e., $A = P(1 + r)^t$

where $p$ – Principal, $A$ - Amount, $r$ - Rate of interest, $t$ -Number of years(period)

Compound Interest (C I) = Amount – Principal

$= P(1 + r)^t - P$

**Illustration 1.1.9**

Find compound interest on ₹20000 at 10% for three years.

**Solution**

P = 20000

r = 10 % = $\frac{10}{100}$

t = 3

$$A = 20000\left(1+\frac{10}{100}\right)^3$$

$$= 20000\left(1+\frac{1}{10}\right)^3$$

$$= 20000 \times \left(\frac{11}{10}\right)^3$$

$$= 26620$$

Compound interest = Amount - Principal

= 26620 − 20000

= ₹6620

**Illustration 1.1.10**

Find compound interest on ₹20000 for 3 years if rate of interest is 5%, 10%, and 20% for first, second and third years respectively.

**Solution**

Amount after 3 years, $A = P(1+r_1)(1+r_2)(1+r_3)$

$$= 20000 \times \left(1+\frac{5}{100}\right)\left(1+\frac{10}{100}\right)\left(1+\frac{20}{100}\right)$$

$$= 20000 \times \left(1+\frac{1}{20}\right)\left(1+\frac{1}{10}\right)\left(1+\frac{1}{5}\right)$$

$$= 20000 \times \frac{21}{20} \times \frac{11}{10} \times \frac{6}{5}$$

$$= ₹\ 27720$$

Compound interest = 27720 – 20000 = ₹7720

**Illustration 1.1.11**

Find compound interest on ₹25000 at 20% p.a for $2\frac{1}{2}$ years, if interest is compounded annually.

**Solution**

$P = 25000$

$r = 20\% = 0.2$

$A = 25000 \times (1 + 0.2)^{2\frac{1}{2}}$

$A = 25000 \times (1.2)^{2\frac{1}{2}}$

$= 25000 \times (1.2)^{2.5}$

$= 25000 \times (1.2)^{2.5}$
$= 39436.02$

Interest = 39436.02 − 25000 = ₹14436.02

**Illustration 1.1.12**

A certain sum will amount to ₹24200 in 2 years at 20% per annum of compound interest, interest being compounded annually. Find the sum.
**Solution**

In this problem we have to find principal.
Given $A$ = 24200, $t$ = 2, $r$ = 20%

$A = P(1 + r)^t$

$$P = \frac{A}{(1 + r)^t}$$

$$= \frac{24200}{\left(1 + \frac{20}{100}\right)^2}$$

$$= \frac{24200}{(1.2)^2}$$

Sum = ₹16805.56

**Illustration 1.1.13**

In how many years will ₹4000 amount to ₹4840 at 10% compound

**Solution**

$A = 4840,\ P = 4000,\ r = \frac{10}{100} = \frac{1}{10}$

$A = P(1 + r)^t$

$4840 = 4000\left(1 + \frac{10}{100}\right)^t$

$\frac{4840}{4000} = \left(1 + \frac{1}{10}\right)^t$

$\frac{121}{100} = \left(\frac{11}{10}\right)^t$

$\left(\frac{11}{10}\right)^2 = \left(\frac{11}{10}\right)^t$

$t = 2\ years$

**Illustration 1.1.14**

At what rate percent per annum of compound interest, compounded annually, will a sum be 16 times of itself in 4 years.

**Solution**

$A = 4P, t = 4$

$A = P(1 + r)^t$

$16P = P(1 + r)^4$

$16 = (1 + r)^4$

$2^4 = (1 + r)^4$

$1 + r = 2$

$r = 1$

$\therefore$ rate of interest $= \frac{r}{100} = 1 \Rightarrow r = 100$

**Illustration 1.1.15**

A certain sum of money double itself in 4 years. In how many years will it become 16 times at the same rate of compound interest.

**Solution**

$A = P(1 + r)^t$

Sum has doubled in 4 years, $A$ becomes $2P$ in 4 years

$$2P = P(1+r)^4$$

$$2 = (1+r)^4$$

$$1 + r = (2)^{\frac{1}{4}}$$

When the sum will become 16 times

$$16P = P(1+r)^t$$

$$16 = (1+r)^t$$

$$(2)^4 = (1+r)^t$$

$$(2)^4 = \left((2)^{\frac{1}{4}}\right)^t$$

$$(2)^4 = (2)^{\frac{t}{4}}$$

$$\frac{t}{4} = 4$$

$t = 16$ years

**Half Yearly Compounding**

When interest is compounded semi-annually (half-yearly), it means that the interest is calculated and added to the principal balance every six months. Here interest is calculated 2 times in a year. Then total number of periods $2t$. To calculate the future value of an investment or loan with half-yearly compounding, you can use the following formula:

$$A = P\left(1 + \frac{r}{2}\right)^{2t}$$

**Quarterly Compounding**

When interest is compounded quarterly, it means that the interest is calculated and added to the principal balance every three months (quarterly). Here interest is calculated 4 times in a year. Then total number of periods $4t$. To calculate the future value of an investment or loan with quarterly compounding, you can use the following formula:

$$A = P\left(1 + \frac{r}{4}\right)^{4t}$$

**Monthly Compounding**

Here interest is calculated 12 times in a year. Then total number of periods $12t$

$$A = P\left(1+\frac{r}{12}\right)^{12t}$$

**Daily Compounding**

Here interest is calculated 365 times in a year. Then total number of periods $365t$

$$A = P\left(1+\frac{r}{365}\right)^{365t}$$

***k* time compounding**

Here interest is calculated k times in a year. Then total number of periods $kt$

$$A = P\left(1+\frac{r}{k}\right)^{kt}$$

**Continuous Compounding (Infinite Compounding)**

Continuous compounding, also known as infinite compounding, is a method of calculating interest or growth where the interest is compounded an infinite number of times over a continuous time period. This is achieved using the mathematical constant "e" (approximately equal to 2.71828 called Euler's number). The formula for calculating the future value of an investment with continuous compounding is as follows:

$$A = P\,e^{rt}$$

where $P$ – Principle, $A$ – Amount, r - annual interest rate, t- time the money is invested for (in years).

**Illustration 1.1.16**

Find the value of ₹100000 at 12% interest rate after 2 years

i. Annual compounding
ii. By-annual compounding
iii. Quarterly compounding
iv. Continuous compounding

i. Maturity value $A = P(1+r)^t$

$$= 100000\left(1+\frac{12}{100}\right)^2$$

$= 100000\,(1.12)^2 = 100000 \times 1.2544$
$= 125440$

ii. Maturity value $A = P\left(1+\frac{r}{2}\right)^{2t}$

$= 100000\left(1+\frac{0.12}{2}\right)^{4}$
$= 100000\,(1+0.06)^4$
$= 100000\,(1.06)^4$
$= 100000 \times 1.262477$
$= 126247.7$

iii. Maturity value $A = P\left(1+\frac{r}{4}\right)^{4t}$

$= 100000\left(1+\frac{0.12}{4}\right)^{8}$
$= 100000\,(1+0.03)^8$
$= 100000\,(1.03)^8$
$= 100000 \times 1.26677$
$= 126677$

iv. Maturity value $A = P\,e^{rt}$

$= 100000 \times e^{0.12\times 2}$
$= 100000 \times 2.71828^{0.24}$
$= 100000 \times 1.271249$
$= 127124.9$

### 1.1.4 Effective Yield

The interest rate that is stated on an annual basis is typically referred to as the nominal interest rate. However, when compounding occurs more frequently than once a year, the actual annualised interest rate tends to be higher than the initially specified nominal interest rate. This higher rate is known as the effective interest rate.

In essence, the effective yield represents the return on an investment where interest payments are reinvested at the same rate. It captures the total yield received by an investor, in contrast to the nominal yield, which doesn't account for the impact of compounding.

If compounding is done $k$ times in a year, then Amount $A = P\left(1+\frac{r}{k}\right)^{kt}$ -------------(1)

If $\hat{r}$ is the effective annualised interest rate then on the basis of annual compounding

Then Amount $A = P(1+\hat{r})^t$------------------- (2)

Now to obtain annualised interest rate we equate (1) and (2)

$$P\left(1+\frac{r}{k}\right)^{kt} = P(1+\hat{r})^t$$

$$\left(1+\frac{r}{k}\right)^{k} = (1+\hat{r})$$

$$\hat{r} = \left(1+\frac{r}{k}\right)^{k} - 1$$

**Illustration 1.1.17**

An investment of ₹100000 grows at 12% per annum, compounded by Half yearly. Calculate the maturity value after 2 years. Also find the effective yield.

**Solution**

Here, $p = ₹100000,\ r = \frac{12}{100} = 0.12,\ \ t = 2, k = 2$

Maturity value $A = P\left(1+\frac{r}{k}\right)^{kt}$

$$= 100000\left(1+\frac{0.12}{2}\right)^{2\times 2}$$
$$= 100000\,(1+0.06)^{4}$$
$$= 100000\,(1.06)^{4}$$
$$= 100000 \times 1.2625$$
$$= 126247.696$$

Effective Yield $\hat{r} = \left(1+\frac{r}{k}\right)^{k} - 1$

$$= \left(1+\frac{0.12}{2}\right)^{2} - 1$$
$$= (1+0.06)^{2} - 1$$
$$= (1.06)^{2} - 1$$
$$= 1.1236 - 1$$
$$= 0.1236$$
$$= 12.36\%$$

**Illustration 1.1.18**

An investment of ₹100000 grows at 12% per annum, compounded quarterly. Calculate the maturity value after 2 years. Also find the effective yield.

**Solution**

Here, $p = ₹100000,\ r = \frac{12}{100} = 0.12,\ \ k = 4,\ t = 2$

Maturity value $A = P\left(1+\frac{r}{k}\right)^{kt}$

$$= 100000\left(1+\frac{0.12}{4}\right)^{4\times2}$$
$$= 100000\,(1+0.03)^{8}$$
$$= 100000\,(1.03)^{8}$$
$$= 100000\ \times 1.26677$$
$$= 126677$$

Effective Yield $\hat{r} = \left(1+\frac{r}{k}\right)^{k} - 1$

$$= \left(1+\frac{0.12}{4}\right)^{4} - 1$$
$$= (1+0.03)^{4} - 1$$
$$= (1.03)^{4} - 1$$
$$= 1.1255 - 1$$
$$= 0.1255$$
$$= 12.55\%$$

**Illustration 1.1.19**

An investment of ₹100000 grows at 12% per annum, compounded monthly. Calculate the maturity value after 2 years. Also find the effective yield.

**Solution**

Here, $P = ₹100000,\ r = \frac{12}{100} = 0.12,\ \ k = 12,\ t = 2$

Maturity value $A = P\left(1+\frac{r}{k}\right)^{kt}$

$$= 100000\left(1+\frac{0.12}{12}\right)^{12\times2}$$
$$= 100000\,(1+0.01)^{24}$$
$$= 100000\,(1.01)^{24}$$
$$= 100000\ \times 1.26973$$
$$= 126973.5$$

Effective Yield $\hat{r} = \left(1+\frac{r}{k}\right)^{k} - 1$

$$= \left(1+\frac{0.12}{12}\right)^{12} - 1$$
$$= (1+0.01)^{12} - 1$$
$$= (1.01)^{12} - 1$$
$$= 1.126825 - 1$$
$$= 0.1268$$
$$= 12.68\ \%$$

## Recap

- Simple Interest = Principal x Rate x Time.
- Compound interest - a method of calculating interest where the interest earned is added to the principal amount, and then future interest is calculated based on the new, larger principal.
- Half Yearly Compounding - interest is calculated and added to the principal balance every six months.
- Quarterly Compounding- interest is calculated and added to the principal balance every three months.
- Monthly Compounding- interest is calculated and added to the principal balance every month.
- Daily Compounding- interest is calculated 365 times in a year.
- Continuous Compounding-Infinite Compounding
- Effective Yield- return on investment that has its interest payments reinvested at the same rate.

## Objective Questions

1. What is the formula for calculating simple interest?
2. If the principal amount is ₹1,000, the annual interest rate is 5%, and the time period is 3 years, what is the simple interest earned?
3. If the principal amount is ₹2,500, the simple interest earned is ₹375, and the time period is 5 years, what is the annual interest rate?
4. What is compound interest?
5. If you borrow ₹5,000 at an annual interest rate of 8% compounded monthly for 2 year, how much will you owe at the end of the 2 years?
6. If you invest ₹1,000 at an annual interest rate of 5% compounded by-annually for 3 years, what will be the effective yield?
7. Continuous compounding is also called
8. In which type of interest, is calculated in 365 terms in a year?

## Answers

1. $I = p\,r\,t$
2. ₹150
3. 3%
4. Interest calculated on both the principal and previously earned interest.
5. 5864.4
6. 5.06 %
7. Infinite compounding
8. Daily compounding

## Self-Assessment Questions

1. The present population of a city is 360000. If it increases at the rate of 5% per annum, find population after 2 years.
2. Find compound interest on ₹15000 at 15 % p.a. for $2\frac{1}{2}$ years.
3. A sum of money invested at compound interest doubles itself in 6 years. At the same rate of interest, it will amount to 8% times of itself in?
4. Discuss the differences between simple interest and compound interest. Explain how these interest types are calculated and their respective applications in the financial world.

## Assignments

1. The simple interest on a sum of money for 2 years is ₹150 and the compound interest on the same sum at same rate for 2 years is ₹155.Find the rate % p.a.

   Ans: 5%

2. Ram borrowed ₹ 8000 at 10% per annum for 2 years. Instead of paying at compound interest, if he pays at simple interest, how much will he gain?
   Ans: ₹1600

3. A sum of ₹1250 is lent for 3 years at 8% per annum of simple interest and compound interest. Find
   a. The amount at the end of first year
   b. The amount at the end of second year
   c. The amount at the end of third year

   Ans: 1) 1350, 1350 2) 1450, 1458 3) 1550, 1574.64

4. The simple interest on a certain sum of money for 2 years at 15% per annum is ₹960. What will be the compound interest on the same sum for the same period at the same rate, interest being calculated yearly?

   Ans: ₹1032

5. The difference in simple interest and compound interest on a certain sum of money in 3 years at 10 % p.a. is ₹372. What is the sum?

   Ans: ₹12000

6. An investor makes an initial investment of ₹10000 for two years. Find the value of the investment after the two years if the investment earns a return of 2 % compounded quarterly. Also, find the effective yield.

   Ans: ₹10407.07

7. If an initial investment of ₹1000 is invested at 8% interest per year, how much would be in the account after five years

   1. with continuous compounding
   2. with quarterly compounding
   3. with annual compounding

   Ans: ₹1491.82, ₹1485.95, ₹1,469.33

8. Roy is looking at investment opportunities and has ₹15000 to invest, with an expected interest of 14% over the next year. Use annual, monthly, quarterly, daily, and continuous compounding to find the ending value of the investment.

   Ans: ₹17100, ₹17204.13, ₹17212.84, ₹17253.64, ₹17254.11

## Suggested Readings


1. Sancheti and V.K Kapoor, (2014). *Business Mathematics*, Sultan Chand & Sons
2. Wikes,F M .(1998). – *Mathematics for Business, Finance and Economics*, Thomson Learning
3. Dr. P.R Vittal. (2012).– *Business Maths & Statistics*, Margham Publications
4. C R Kothari. (2013).– *Quantitative Techniques*, Vikas Publishing House
5. Goel and Goel. (2014).- *Mathematics and Statistics*, Taxmann Allied Services Pvt Ltd

# Unit - 2
# Time Values

## Learning Outcomes

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

- ✓ comprehend the concept of present value and future value and their importance in finance
- ✓ calculate the present value of a future cash flow or investment using the appropriate discount rate.
- ✓ apply present value calculations to make financial decisions, such as investment analysis and valuation.
- ✓ grasp the concept of depreciation and amortization

## Prerequisites

In a quiet town, there lived a wise old man named Arjun who loved to teach life lessons through stories. One day, he called his two young friends, Meera and Raj, and handed each of them a shiny gold coin.

"Here's a challenge," Arjun said. "You have two options: I can give you **one gold coin today**, or I can give you **two gold coins after one year.** What will you choose?"

Meera thought for a moment and said, "I'll take the two gold coins after a year because more is better!"

Raj, on the other hand, said, "I'll take the one gold coin today."

Arjun smiled and asked Raj, "Why would you choose one coin today when you could get two later?"

Raj replied, "Because if I have the coin today, I can use it to earn more. For example, I could lend it to someone and earn interest, or I could buy seeds to grow crops and make a profit. Waiting a whole year means losing the chance to make money now."

Meera looked surprised. “I didn’t think of that!

Arjun nodded. “Raj is right. Money today is worth more than the same amount in the future because of its **earning potential**. This is the **time value of money**. A gold coin in your hand today can be invested or used to create more wealth.”

Meera smiled and said, “I’ll remember this lesson! Money has more value when you use it wisely.”

The two friends thanked Arjun and left, wiser about the importance of the time value of money.

## Keywords

Present value, Future value, Annuity, Depreciation, Amortisation

## Discussion

### 1.2.1 Introduction

Time value is a fundamental concept in finance and economics, representing the notion that the value of money changes over time due to various factors, including inflation, interest rates, and opportunity costs. Essentially, it asserts that the worth of a sum of money today is not equivalent to its worth in the future. The principle underlying time value is that money available today can be invested or used to generate returns over a given period. As a result, the same amount of money in the future would be worth less due to the potential returns it could have generated if invested today. Conversely, the value of money is viewed as more valuable in the present since it can be invested to accrue additional value over time.

This concept profoundly influences decision-making in finance, investments, and business. Individuals, companies, and governments use time value to evaluate investment opportunities, determine loan interest rates, plan for retirement, assess the profitability of projects, and make informed financial choices. Understanding the time value of money enables better financial planning by assessing the implications of various financial decisions over different time horizons. It guides investors to consider risk, potential returns, and inflation when making investment choices, ensuring prudent financial management and sustainable growth. In fact, the time value of money is a critical concept that shapes financial strategies and actions, highlighting the importance of considering the temporal aspect in economic and financial decision-making.

## 1.2.2 Present value

Present value (PV) is a financial concept that represents the current worth or value of a sum of money or a series of future cash flows, discounted back to the present moment. In essence, it's the process of determining what a future payment or series of payments is worth today, given the time value of money.

The underlying idea of present value is based on the principle that money received or spent in the future is not equivalent in value to the same amount of money received or spent today. This is primarily due to factors such as inflation, investment opportunities, and risk. Therefore, to make informed financial decisions, it's important to calculate the present value of future cash flows.

The formula for calculating the present value of a single future payment is as follows:

Present Value PV = $FV \times \left(1+\frac{r}{n}\right)^{-nt}$

where, $FV$ is the Future Value,
$r$ is the interest rate,
$n$ is the number of times that interest is compounded per year. For e.g.; If the interest is calculated quarterly, 'n' would be four.
$t$ is the number of years the money is invested or saved for.

Under continuous compounding the present value can be calculated by using the following formula:

Present Value (PV) = FV$\times$ $(e)^{-rt}$
Where, e is a mathematical constant approximately equal to 2.71828.
r is the interest rate
t is the number of years the money is invested

The concept of present value is widely used in various financial applications, including investment analysis, loan valuation, retirement planning, and capital budgeting. It allows individuals and businesses to make informed decisions by comparing the current value of future cash flows to their current financial situation and investment opportunities.

### 1.2.3 Future value

Future value (FV) is a financial concept that represents the worth or value of a sum of money at a specified point in the future, based on an assumed rate of return or interest rate. It's a crucial concept in finance, investment, and personal financial planning, as it helps individuals and businesses project the growth of their investments or savings over time.

The future value is calculated by taking into account the initial principal amount (the amount of money you have now), the rate of return (usually expressed as an annual interest rate), and the time period for which the money is invested or saved.

The formula to calculate the future value of an investment or savings with compound interest is as follows:

$$\text{Future Value FV} = PV \times \left(1+\frac{r}{n}\right)^{nt}$$

$$\text{Future Value} = \text{Present Value } \left(1+\frac{r}{n}\right)^{nt}$$

Where,

$PV$ is the present value

r is the annual interest rate.

$n$ is the number of times that interest is compounded per year.

$t$ is the number of years the money is invested or saved for.

In simpler terms, the future value tells you how much a sum of money will grow over time when it earns interest or experiences investment returns, assuming that those returns are reinvested or compounded.

For example, if you invest ₹1000 in a savings account with an annual interest rate of 5%, the future value of your investment after 5 years would be calculated using the formula above. This calculation helps you estimate how much money you will have in the account at the end of the 5 year period.

**Illustration 1.2.1**

Calculate the present value for an income of ₹1000000 after 5 years given the interest rate of 9% per annum

(1) Under annual compounding

(2) Under quarterly compounding

(3) Under continuous compounding

**Solution**

(1) Under annual compounding

Present Value PV = $FV \times \left(1+\frac{r}{n}\right)^{-nt}$

$$= 1000000 \times \left(1+\frac{.09}{1}\right)^{-5}$$

$$= \frac{1000000}{(1.09)^5}$$

$$= \frac{1000000}{(1.09)^5}$$

$$= \frac{1000000}{1.5386}$$

$$= 649931.4$$

2) Under quarterly compounding

Present Value PV = $FV \times \left(1+\frac{r}{n}\right)^{-nt}$

$$= 1000000 \times \left(1+\frac{.09}{4}\right)^{-4\times5}$$

$$= \frac{100,0000}{\left(1+\frac{0.09}{4}\right)^{4\times5}}$$

$$= \frac{1000000}{(1.0225)^{20}}$$

$$= \frac{1000000}{1.56051}$$

$$= 640816.14$$

3) Under continuous compounding

Present Value PV = $FV \times (e)^{-rt}$

$$= \frac{1000000}{(2.71828)^{0.09\times5}}$$

$$= \frac{1000000}{(2.71828)^{0.45}}$$

$$= \frac{1000000}{1.568311}$$

$$= 637628.63$$

**Illustration 1.2.2**

Rani has ₹9000 in her bank account and she earns an annual interest 4%. Find the future value of her account after 5 years.

(1) Under annual compounding
(2) Under quarterly compounding
(3) Under continuous compounding

**Solution**

1) Future Value FV = $PV \times \left(1+\frac{r}{n}\right)^{nt}$

$= 9000 \times \left(1+\frac{0.04}{1}\right)^{5}$

$= 9000 \times (1.04)^{5}$

$= 10949.88$

2) Future Value FV = $PV \times \left(1+\frac{r}{n}\right)^{nt}$

$= 9000 \times (1+0.01)^{20}$

$= 9000 \times (1.01)^{20}$

$= 10981.71$

3) Future Value FV = $PV \times (e)^{rt}$

$= 9000 \times (2.71828)^{0.04\times 5}$

$= 9000 \times 1.2214$

$= 10992.63$

## 1.2.4 Annuity

An annuity is a financial product that provides a fixed amount of money at regular intervals for a specified period. Annuities are primarily purchased by individuals seeking a stable, regular income. Essentially, an annuity is a series of equal installments that guarantee income to an individual.

There are several practical situations involving financial transactions with annuities:

- Repaying a home loan with equal monthly installments.
- A retired person purchasing an annuity from an insurance company upon retirement.
- Purchasing a life insurance policy with monthly premiums.

In these scenarios, an individual initially pays a lump sum to a financial institution, which then holds the amount for a set period. Afterward, the financial institution disburses fixed payments to the individual at predetermined intervals.

### 1.2.4.1 Types of Annuities

There are several types of annuities according to frequency and types of payments. The various types of annuities are:

**1. Annuity certain**

Here the term of an annuity begins and ends on a certain fixed date. An annuity certain, also known simply as an "annuity," is a financial arrangement that involves a series of periodic payments or receipts of a fixed amount over a predetermined period. The defining characteristic of an annuity certain is that the number of payment periods is specified and certain from the outset. These payments can be made at regular intervals, such as monthly, quarterly, or annually, and can be received or paid out for a specified number of periods.

An annuity certain is a useful financial tool for individuals and organizations to plan for a predictable stream of payments or receipts over a specified period, ensuring financial stability and security during that timeframe.

**2. Perpetuity**

An annuity that provides cash flows without an end date. In this type of annuity, only interest payments are made regularly, with no repayment of the principal. As a result, the annuity never terminates. A perpetuity begins on a specific date but continues indefinitely until the entire principal amount is repaid.

**3.Contingent annuity.**

Contingent annuity begins only a particular date and terminates at a future happening Example: The payment of LIC premium starts on a particular date but ends when concerned person dies.

**4.Ordinary Annuity (Annuity Immediate):**

An ordinary annuity, also known as a regular annuity, is a financial arrangement in which a series of equal payments or receipts are made or received at the end of each payment period. In other words, in an ordinary annuity, the payments occur at the end of each time period, whether it's monthly, quarterly, annually, or some other defined interval. Payment is made at the beginning of each period is called Annuity Due.

## 1.2.5 Valuation of Annuities

Annuities are valued by discounting the future cash flows of the annuities and finding the present value of the cash flows. The formula for annuity valuation is:

$$\text{Present Value } PV = PMT\left(\frac{1-(1+r)^{-n}}{r}\right)$$

$$\text{Future Value } FV = PMT\left(\frac{(1+r)^{n}-1}{r}\right)$$

Where:

- PV = Present value of the annuity
- FV = Future value of the annuity
- PMT = Periodic payment amount
- r = Interest rate
- n = Total number of periods of annuity payments

**Illustration 1.2.3**

A person pays ₹ 640000 per annum for 12 years at the rate of 10% per year. Find the annuity.

**Solution**

Future Value $FV = PMT\left(\frac{(1+r)^n - 1}{r}\right)$

$$= 640000\left(\frac{(1+0.1)^{12} - 1}{0.1}\right)$$

$$= 640000\left(\frac{(1.1)^{12} - 1}{0.1}\right)$$

$$= 640000\left(\frac{3.13843\text{-}1)}{0.1}\right)$$

$$= 640000\left(\frac{2.13843}{0.1}\right)$$

$$= 640000 \times 21.384$$

$$=13685760$$

**Illustration 1.2.4**

What amount should be deposited annually so that after 10 years a person receives ₹150000 if the interest rate is 15%.

**Solution**

$FV = 150000, r = 0.15,\ \ n = 10$

$$FV = PMT\left(\frac{(1+r)^n - 1}{r}\right)$$

$$150000 = PMT\left(\frac{(1+0.15)^{10} - 1}{0.15}\right)$$

$$150000 = PMT\left(\frac{4.046 - 1}{0.15}\right)$$

$$150000 = PMT\left(\frac{3.046}{0.15}\right)$$

$$PMT = 150000 \times \left(\frac{0.15}{3.046}\right)$$

$$= 7386.76$$

## 1.2.6 Depreciation

Depreciation is the process of allocating the cost of a fixed asset over different accounting periods to reflect the benefit derived from its use. This concept applies to all assets whose benefits extend over a long period, typically more than one year. Examples include machinery, furniture, buildings, and leases. Therefore, depreciation plays a crucial role in income measurement.

Depreciation represents a permanent decline in the value of an asset. Over time, all fixed assets are expected to become less efficient. Depreciation is calculated as an estimate of this wear and tear and is recorded in the Profit & Loss account, either on a monthly or annual basis. The Net Book Value of the asset is determined by subtracting the total depreciation from its original cost

According to Spicer and Pegler "Depreciation is the measure of the exhaustion of the effective life of an asset from any cause during a given period."

**Important Methods of Calculating Depreciation**

1. Straight Line Method or Fixed Instalment Method.
2. Written Down Value Method or Diminishing Balance Method.

**1. Straight-line Method or Fixed Instalment Method or Original Cost Method**

The Straight-Line Method, also known as the Fixed Instalment Method, is an accounting approach used to calculate depreciation by evenly distributing the cost of a fixed asset over its estimated useful life, resulting in a constant depreciation expense each accounting period.

The formula for the Straight-Line Method is

$$\text{Depreciation Expense} = \frac{\text{Cost of Asset} - \text{Salvage Value}}{\text{Estimated Useful Life}}$$

where:

- Depreciation Expense is the amount recognized as an expense on the income statement.
- Cost of Asset is the initial purchase price or acquisition cost of the asset.
- Salvage Value is the estimated residual or salvage value of the asset at the end of its useful life (sometimes referred to as the "scrap value" or "residual value").
- Estimated Useful Life is the expected number of accounting periods or years over which the asset is expected to provide benefits or generate revenue.

This method is straightforward and provides a consistent, linear allocation of an asset's cost over its useful life, making it easy to apply and understand for financial reporting and tax purposes.

Under this method, the same amount of depreciation is charged every year throughout the life of the asset.

**Advantages**

**i) Simplicity and ease of use**: The straight-line method is straightforward and easy to understand. It involves a simple calculation where the cost of the asset is divided equally over its useful life. This simplicity makes it widely accessible and applicable for businesses of all sizes and industries.

**ii) Predictability and consistency**: The straight-line method provides a consistent and predictable pattern of depreciation. Since the depreciation expense remains the same throughout the asset's useful life, it allows for easier financial planning and budgeting. This predictability aids in more accurate financial forecasting and decision making.

**iii) Equal allocation of costs:** By evenly spreading the cost of the asset over its useful life, the straight-line method ensures that the expense is allocated fairly and consistently. This aligns with the matching principle in accounting, where expenses are matched with the revenues they generate.

**Disadvantages**

**i) Unrealistic assumption**: The straight-line method assumes that the asset depreciates evenly over time. However, in reality, many assets may have different patterns of value decline. Some assets might experience higher depreciation in the early years, while others may have a more rapid decline in later years. The straight-line method does not account for these variations.

**ii) Inaccurate reflection of asset value**: Due to its linear nature, the straight-line method may not accurately represent the asset's actual value at any given point in time. The asset may be worth significantly more or less than its recorded value on the Balance Sheet. This discrepancy can affect the financial reporting and decision making process.

**iii)Tax implications**: Depending on the tax regulations in a particular jurisdiction, using the straight-line method may not provide the most advantageous tax benefits. Other depreciation methods, such as accelerated depreciation, may allow for larger deductions in the early years, resulting in reduced tax liabilities.

**iv) Ignoring technological advancements:** In industries with rapid technological advancements, assets can quickly become outdated or obsolete. The straight-line method does not account for this factor and may result in overestimating the asset's useful life and undervaluing the need for asset replacement or upgrades.

**2. Written Down Value Method or Diminishing Balance Method or Reducing Balance Method**

The Written Down Value Method, also known as the Diminishing Balance Method or Reducing Balance Method, is an accounting approach used to calculate depreciation by applying a fixed percentage rate to the diminishing book value of a fixed asset, resulting in a decreasing depreciation expense each accounting period.

Under this method, depreciation is charged at a fixed percentage each year on the reducing balance (i.e., cost less depreciation) goes on decreasing every year.

**Advantages**

**i) Reflects asset's actual usage:** The diminishing balance method aligns with the concept that assets often experience a higher rate of depreciation in their early years due to heavy usage or technological advancements. This method better reflects the asset's actual decline in value over time, especially for assets that quickly become outdated or less efficient.

**ii) Tax benefits:** The accelerated depreciation provided by the diminishing balance method can result in higher depreciation deductions in the earlier years. This can lead to increased tax benefits and reduced tax liabilities for businesses, providing more immediate cash flow advantages.

**iii) Early expense recognition:** The diminishing balance method allows businesses to front-load depreciation expenses. This means that a larger portion of the asset's cost is expensed in the earlier years, which can provide a more accurate reflection of the asset's contribution to revenue generation during its more productive stages.

**Disadvantages**

**i) Overstated asset value in later years:** As the diminishing balance method applies a higher depreciation rate in the early years, the asset's net book value may be overstated in the later years of its useful life. This can lead to an inflated representation of the asset's value on the Balance Sheet, potentially misrepresenting its true economic worth.

**ii) Inconsistent expense recognition:** Unlike the straight-line method, where depreciation expense is evenly distributed, the diminishing balance method results in varying depreciation expenses each period. This can make financial analysis and year-over-year comparisons more challenging, as the depreciation expenses fluctuate based on the asset's net book value.

**iii) Limited applicability:** The diminishing balance method may not be suitable for all types of assets. It is typically more appropriate for assets with a higher rate of obsolescence or assets which rapidly decline in value. Assets that have a more consistent or linear decline may be better served by other depreciation methods, such as straight-line depreciation.

**iv) Value not reduced to zero:** The disadvantage of the diminishing balance method is that it does not reduce the asset's value to zero within its useful life. Unlike some other depreciation methods, such as the straight-line method, the diminishing balance method may not fully write off the asset's value by the end of its useful life.

**Illustration 1.2.5**

A Car is purchased for ₹ 10 lakhs Calculate Depreciation as per

1)Straight line method
2)Written down value method @15% per annum for the first 5 years.

**Solution**

1) Straight line method

Depreciation for each year =1000000 × $\frac{15}{100}$ = ₹150000

| Year | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Depreciation (₹) | 150000 | 150000 | 150000 | 150000 | 150000 |

(ii) Written down value method (WDV)

| Year | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| WDV at beginning (Rs) | 1000000 | 850000 | 722500 | 614125 | 522006.25 |
| Depreciation (Rs) | 150000 | 127500 | 108375 | 92118.75 | 78300.94 |
| WDV at end (Rs) | 850000 | 722500 | 614125 | 522006.25 | 443705.31 |

## 1.2.7 Amortisation

Amortization is the process of spreading the cost of an asset over its useful life. It is used in both accounting and finance.

In accounting, amortization is used to reduce the book value of an intangible asset over time. Intangible assets are assets that do not have a physical form, such as patents, trademarks, and copyrights. They have a limited useful life, and their value declines over time. Amortization is used to match the expense of the asset to the revenue it generates.

Example: A company buys a patent for ₹1000000. The patent has a useful life of 10 years. The company would amortize the patent over 10 years, meaning that it would expense ₹100000 per year for the patent.

In finance, amortization is used to calculate the monthly payments on a loan. The loan balance is amortized over the term of the loan, meaning that a portion of each payment goes towards paying down the principal and the rest goes towards paying interest. Over time, a larger portion of each payment goes towards principal, and a smaller portion goes towards interest.

Example: A person takes out a ₹300000 mortgage with a 30 year term and a 5% interest rate. The monthly payments on the mortgage would amortize the loan balance over 30 years. This means that a portion of each payment would go towards paying down the principal and the rest would go towards paying interest. Over time, a larger portion of each payment would go towards principal, and a smaller portion would go towards interest.

## Recap

- Time value- money available today is worth more than the same amount of money in the future.
- Annuity- a financial arrangement that provides a series of periodic payments at regular intervals, over a specified time period.
- Depreciation-permanent and continuing diminution in the quality, quantity or the value of an asset.
- Straight line methods- amount of depreciation is same ever year.
- Diminishing balance method- amount of depreciation decreases each year.
- Scrap value- residual value.
- Amortisation- process of systematically reducing or paying off a debt, typically a loan or mortgage, through scheduled payments that include both principal and interest over a specified period.

## Objective Questions

1. What do we call a series of payments that follow a regular and fixed sequence for seeking a regular income?
2. What is the time between successive payment of an annuity?
3. What is the process of systematically reducing a loan, through scheduled payments that include both principal and interest over a specified period.
4. Which method is used to allocate the cost of a tangible asset over its estimated useful life, reflecting the asset's declining value over time?
5. In which type of annuity do payments starts immediately at the end of each period?
6. Money available today is worth more than the same amount of money in future means what concept?
7. In which method amount of depreciation decrease in every year?
8. Residual value is also called?

## Answers

1. Annuity
2. Payment Period
3. Amortisation
4. Depreciation
5. Annuity Due
6. Time value of money
7. Diminishing Balance Method
8. Scrap value

## Self-Assessment Questions

1. Explain the concept of time value of money and its significance in financial decision- making. Provide examples to illustrate its application.
2. Discuss the various factors that influence the time value of money. How do these factors impact investment decisions and financial planning?
3. You invest ₹5000 in an account that earns an annual interest rate of 6%. How much will you have after 3 years?
4. You want to have ₹100000 in 5 years for a down payment on a house. If you can earn an annual interest rate of 4%, how much do you need to invest today to reach your goal?

## Assignments

1. What is the present value of ₹10000 received in two years if the interest rate is
    a. 12% per year discounted annually
    b. 12% per year discounted semi-annually
    c. 12% per year quarterly

    **Ans: ₹7971.94, ₹7920.9, ₹7894.52**

2. How much is ₹7000 worth today if received 10 years from now with a 7% interest rate?

    **Ans: ₹3558.45**

3. If we want ₹ 2000 three years from now and the compounded interest rate is 8%, how much should we invest today?

    **Ans: ₹1587.66**

4. How much would you have to deposit today to have ₹10000 in five years at 6% interest discounted quarterly?

    **Ans: ₹7424.72**

5. What is the present value of an offer of 15000 one year from now if the opportunity cost of capital (discount rate) is 12% per year nominal annual rate compounded monthly?

    **Ans: ₹13311.74**

## Suggested Readings


1. Allen, R.G.D. (1949). *Statistics for Economics*, Hutchinson & Co. (Publishers) Ltd., London.
2. Charles, A. Decision-making under Uncertainty: Models and Choices, Prentice Hall, Englewood Cliffs.
3. C R Kothari (2013)– *Quantitative Techniques*, Vikas Publishing House
4. Goel and Goel (2014)- *Mathematics and Statistics*, Taxmann Allied Services Pvt Ltd

# Unit - 3

# Ratios and Proportions

## Learning Outcomes

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

- ✓ comprehend the concept of ratio and proportions
- ✓ get an awareness in the relationship between ratio and proportion
- ✓ applying ratios and proportions to solve real-world problems

## Prerequisites

One sunny afternoon, two friends, Aanya and Riya, decided to make lemonade to sell at the park. Aanya had a great recipe:

"For every **2 cups of lemon juice**, we need **3 cups of water** and **1 cup of sugar**."

Riya was curious. "Why do we need to follow this exact recipe?"

Aanya explained, "If we don't keep the same **ratio**, the lemonade might taste too sour, too sweet, or too watery. The ratio is **2:3:1**, and it helps us make the perfect lemonade every time."

They made one batch of lemonade and sold it quickly. Soon, they decided to make more.
"How much do we need if we want to make **double** the amount?" Riya asked.
Aanya smiled. "We use **proportion** to scale up! If the original ratio is **2:3:1**, we multiply each number by 2. That means we need:

- **4 cups of lemon juice (2 × 2)**
- **6 cups of water (3 × 2)**
- **2 cups of sugar (1 × 2)**."

They followed the proportions and made a bigger batch of lemonade, which was just as delicious as the first.

Riya was amazed. "Ratios and proportions are so useful! Now I understand how they help us keep things consistent."

From that day on, Riya started noticing ratios and proportions everywhere - in cooking, in splitting costs, and even in planning her study schedule!

## Keywords

Ratio, Proposition, Mean proportion, Direct proportion, Indirect proportion

## Discussion

### 1.3.1 Introduction

#### Ratio

A ratio is a way of comparing two or more quantities or values. It expresses the relationship between these quantities in terms of how many times one quantity is contained within another. Ratios are often written in the form of a fraction or using a colon (":") such as 2/3 or 2:3. The first term in a ratio is called antecedent and the second term is called consequent. Ratio has no units. We can multiply and divide the terms of the ratio by a non-zero number.

For example, if you have a basket with 2 apples and 3 oranges, the ratio of apples to oranges is 2:3, which means there are 2 apples for every 3 oranges.

### 1.3.2 Types of Ratios

**Simple Ratio**

A comparison of two quantities without specifying a particular unit of measurement is a simple ratio.

For example, 2:3 is a simple ratio.

**Compound Ratio**

The concept of compound ratios involves combining two or more simple ratios into a single ratio, which can be expressed as a product of these individual ratios.

The compound Ratio of $a:b, c:d, e:f$ can be written in fraction as $\frac{a}{b} \times \frac{c}{d} \times \frac{e}{f}$ which is essentially the result of multiplying these simple ratios together.
For example, 2:3 and 4:5 combined as (2:3): (4:5).

**Continued Ratio**

A continued ratio, also known as a continued fraction, is a mathematical expression that represents a number as an infinite sequence of fractions.

For example, 2:3:4:5…… is a continued ratio

**Equivalent Ratios**

Ratios that represent the same relationship between quantities are considered equivalent. For example, 2:3 is equivalent to 4:6 because they both represent the ratio of two parts to three parts.

### 1.3.3 Uses of Ratios

- Ratios are used to set up proportions, which help solve various mathematical and real-world problems.

- Ratios are used in scaling up or down measurements. For instance, if a recipe serves 4 people and you need to serve 8, you can use a ratio to scale the ingredients accordingly.

- Financial Analysis- Ratios are vital in financial analysis, such as calculating the debt-to-equity ratio or the price-to-earnings ratio.

- Ratios are commonly used in geometry to express the relationship between the sides of similar triangles. This is known as the "similarity ratio."

- Ratios are used in various real-life scenarios, such as cooking (recipe adjustments), construction (scaling building plans), and finance (investment analysis).

**Illustration 1.3.1**

Find the ratio of 650 grams and 26 kg.

**Solution**

Since the ratio is always between the quantities of the same kind, we must convert them into same kind.

Since 26 kg =26000 gm, required ratio is

$$650:26000 = \frac{650}{26000} = \frac{1}{40} = 1:40$$

**Illustration 1.3.2**

Find the ratio of 52 to 117.

**Solution**

$$52:117 = \frac{52}{117} = \frac{13 \times 4}{13 \times 9} = \frac{4}{9} = 4:9$$

**Illustration 1.3.3**

Find the ratio of ₹ 8 to 80 paise.

**Solution**

₹ 8 is 800 paise So, the ratio is 800:80. We can divide the terms by 80. So, the ratio is 10:1

**Illustration 1.3.4**

Find the ratio of 8 gross and 4 dozen.

**Solution**

Since the ratio is always between the quantities of the same kind, we must convert them into same kind.

1 gross =12 dozens,

8 grosses = 8 $\times$ 12 dozens = 96 dozens

8 grosses:4 dozens = 96 dozens:4 dozens = $96:4 = \frac{96}{4} = \frac{24}{1} = 24:1$

**Illustration 1.3.5**

In a mixture of gold and copper are in the ratio 7:3. How much are each in a mixture of 120 grams?

**Solution**

Quantity of Gold in the mixture = $\frac{7}{(7+3)} \times 120 = \frac{7}{10} \times 120 = 84\ grams$

Quantity of Copper in the mixture = $\frac{3}{(7+3)} \times 120 = \frac{3}{10} \times 120 = 36\ grams$

**Illustration 1.3.6**

Income of Rahim is ₹12000 per month and that of Amit is ₹191520 per annum. If the monthly expenditure of each of them is ₹.9960 per month find the ratio of their savings.

**Solution**

Savings of Rahim per month = ₹ $(12000 - 9960) =$ Rs $2040$.

Monthly income of Amit = ₹ $\frac{191520}{12}$ $=$ ₹ $15960$

Savings of Amit per month $=$ ₹$(15960 - 9960) =$ ₹$6000$

Therefore, ratio of savings of Rahim and Amit = 2040: 6000 = 17:50

**Illustration 1.3.7**

If a bus travels 160 km in 4 hours and a train travels 320 km in 5 hours at uniform speeds, then find the ratio of the distances travelled by them in one hour?

**Solution**

Ratio of bus speed: train speed $= \frac{160}{4} : \frac{320}{5} = 40:64 = 5:8$ .

**Illustration 1.3.8**

Find the greatest ratio among the ratios $2:3,\ 5:8,\ 75:12$ and $40:25$ .

**Solution**

$2:3, 5:8,\ 75:12,\ 40:25 = \frac{2}{3}, \frac{5}{8}, \frac{75}{12}, \frac{40}{25}$

L C M = 600

$\frac{2}{3} \times 600 = 400, \frac{5}{8} \times 600 = 375, \frac{75}{12} \times 600 = 3750, \frac{40}{25} \times 600 = 960$

In ascending order, we get 375,400,960, writing the corresponding original ratios

$\frac{5}{8}, \frac{2}{3}, \frac{40}{25}, \frac{75}{12}$

Ratios in the ascending order are 5:8, 2: 3, 40:25, 75:12.

**Illustration 1.3.9**

Find the compound Ratio of $2:3,\ 4:9, 10:13$ is $\frac{a}{b} \times \frac{c}{d} \times \frac{e}{f}$

**Solution**

Compound Ratio = $2:3, 4:9, 10:13 = \frac{2}{3} \times \frac{4}{9} \times \frac{10}{13} = \frac{80}{351}$ = 80:351

**Illustration 1.3.10**

Ramesh and Mahesh have invested in the ratio of 4 : 7. If both invested an amount of ₹ 49500, then find the investment of Mahesh.

**Solution**

The given ratio of investment for Ramesh to Mahesh is 4:7. This means the total ratio is 4 + 7 = 11.

1 part of the ratio = $\frac{\text{Total amount}}{\text{Total ratio}}$ .

1 part = $\frac{49500}{11}$

Mahesh's investment = 7 parts × (Value of 1 part)

$= 7 \times \frac{49500}{11}$

= ₹ 31500

So, Mahesh's investment is approximately ₹ 31500.

**Illustration 1.3.11**

The ratio between present ages of Mohan and Vivek are 5:8. After 4 years, the ratio between their ages will be 2:3. What is the present age of Vivek?

**Solution:**

Let present ages of Mohan and Vivek are 5x and 8x respectively

Given $\frac{5x+4}{8x+4} = \frac{2}{3}$ cross multiplying

$15x + 12 = 16x + 8$

$16x + 8 - 15x - 12 = 0$

$x - 4 = 0$

$\therefore x = 4$

Vivek present ages $= 8x = 8 \times 4 = 32$ years.

## 1.3.4 Proportions

Proportions are mathematical expressions that establish the equality of two ratios. They are used to solve a wide range of problems involving unknown quantities and are a fundamental concept in mathematics.

A proportion is an equation that states that two ratios are equal. It is written in the form of $\frac{a}{b} = \frac{c}{d}$ where a, b, c and d are numbers (or expressions), and the two ratios to be in proportion.

If the numbers $a, b, c,$ and $d$ are in proportion we can express $a\colon b = c\colon d$. It can also be expressed as $a\colon b :: c\colon d$.

In any proportion, the first and fourth terms are called the extremes while the second and the third terms are called means. In the proportion a : $b :: c\colon d$ , $a$ and $d$ are extremes and $b$ and $c$ are means.

Product of extremes = Product of means

i.e., if $a\colon b = c\colon d, then\ a\ x\ d = b\ x\ c$

To solve a proportion, you can use the "cross-multiplication" method. It involves multiplying the extremes (the first and last terms) and the means (the second and third terms) and setting them equal to each other.

For example, Solve for x in the proportion $3\colon 5 :: x\colon 15$

$$3\colon 5 :: x\colon 15 \rightarrow \frac{3}{5} = \frac{x}{15}$$

Cross-multiplying: $3 \times 15 = x \times 5$

$$45 = x \times 5$$

$$x = \frac{45}{5} = 9$$

Proportions are widely used in real-life scenarios such as adjusting ingredient quantities when changing recipe serving sizes, calculating distances on maps and scales, calculating interest rates, exchange rates, and investment returns, finding similar shapes and scaling figures, solving problems involving speed, distance, and time.

### 1.3.5 Continued proportion

If the means of a proportion are equal, then proportion is called continued proportion.

i.e., if $x\colon y :: y\colon z$ then $x, y\ and\ z$ are said to be in continued proportion.

For example: $4\colon 6 :: 6\colon 9$ is a continued proportion.

### 1.3.6 Mean proportion (Second Proportion)

If $a\colon x = x\colon b$ then $x$ is called Second Proportion (Mean proportion) of $a$ and $b$

$$a\colon x = x\colon b \Longrightarrow x^2 = ab \Longrightarrow x = \sqrt{ab}$$

### 1.3.7 Third proportion

If $a : b \ = \ b : x$ then $x$ is called Third proportion of $a$ and $b$

$$a : b \ = \ b : x \Longrightarrow \ ax \ = \ b^2 \ \ \Longrightarrow \ x \ = \frac{b^2}{a}$$

For example, third proportion of 4 and 6. Let x be the third proportion of 4 and 6

By definition $4 : 6 \ = \ 6 : x \ \Longrightarrow 4x \ = 6 * 6 \Longrightarrow \ x = \frac{36}{4} = 9$

### 1.3.8 Fourth proportion

If $a : b \ = \ c : x$ then $x$ is called Fourth proportion of $a, b \ and \ c$.

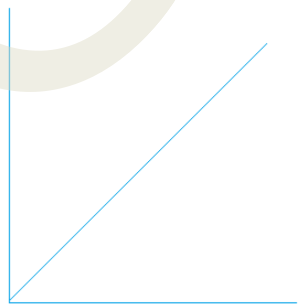
For example, fourth proportion of 6,9 and 20. Let x be the Fourth proportion of 6,9 and 20

By definition $6 : 9 \ = \ 20 : x \ \Longrightarrow 6\,x \ = \ 9 * 20 \Longrightarrow \ x = \frac{180}{6} \Longrightarrow 30$

### 1.3.9 Direct proportion

Two quantities are said to be in direct proportion if increase in one quantity follows the increase in other quantity or decrease in one quantity follows the decrease in other quantity in the same ratio.

Example: Number of articles bought and total cost of article. That is more articles, more cost and less articles, less cost.



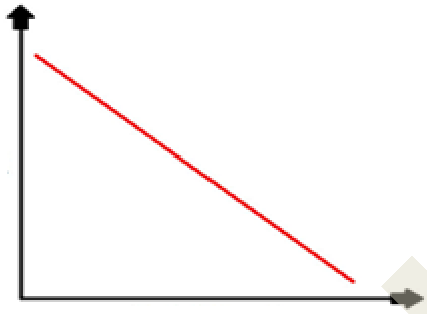
**Direct proportion**

### 1.3.10 Indirect proportion

Two quantities are said to be in indirect or inverse proportion if increase in one quantity follows decrease in the other quantity in same ratio and vice versa.

Example: Number of workers and days taken to do a certain work. i.e., more workers, less days and less days, more workers.

If the two quantities $x$,y are indirectly proportional to each other, then $x = \frac{k}{y}$ or $\frac{x_1}{y_1} = \frac{x_2}{y_2}$



**Indirect proportion**

**Illustration 1.3.12**

What must be added to each of the numbers 10,18,22,38 so that they become in proportion.

**Solution**

Let $x$ be the number to be added to each of the numbers.

$$(10 + x):(18 + x)::(22 + x):(38 + x)$$
$$(10 + x)(38 + x) = (18 + x)(22 + x)$$
$$380 + 48x + x^2 = 396 + 40x + x^2$$
$$48x - 40x = 396 - 380$$
$$8x = 16$$
$$x = 2$$

2 should be added to each of the given numbers

**Illustration 1.3.13**

A man went to the market to buy fruits. He bought apples, mangoes and dates. The ratio of the weights of apples to that of mangoes bought is same as the weight of mangoes to that of dates bought. If he bought 7.2 kg of apples and 5 kg of dates, find the weight of mangoes bought.

**Solution**

Given Apples: Mangoes = Mangoes: Dates

Let $x$ be the weight of mangoes

Apples: $x$ = $x$ : dates

$x^2 = Apples \times dates = 7.2 \times 5 = 36 \Longrightarrow x = 6$

Weight of Mangoes = 6 Kg

**Illustration 1.3.14**

The ratio between two quantities is 9:8. If the first quantity is 45. Find the other quantity.

**Solution**

Suppose the second quantity be $x$, then $9:8 = 45:x$

In a proportion, product of the means = Product of extremes

$$9 \times x = 45 \times 8 \Longrightarrow x = 45 \times \frac{8}{9} = 40$$

**Illustration 1.3.15**

If $a : b = 2:3$, $b:c = 4:5$, $c:d = 6:7$, find $a:d$

**Solution**

$$\frac{a}{b} \times \frac{b}{c} \times \frac{c}{d} = \frac{2}{3} \times \frac{4}{5} \times \frac{6}{7} \Longrightarrow \frac{a}{d} = \frac{48}{105} = \frac{16}{35} \Longrightarrow a:d = 16:35$$

**Illustration 1.3.16**

If $a;b = 2:3$, $b:c = 4:5$, $c:d = 6:7$ find $a:b:c:d$

**Solution**

$a:b = 2:3$ $------$ (1) , $b:c = 4:5$ $--------$ (2)

Here values of $b$ in both ratios is different. So, make both values of $b$ equal.

Multiply (1) by 4 and (2) by 3 we get

$a:b = 8:12$ and $b:c = 12:15$

$i.e., a:b:c = 8:12:15$ --------------(3)

$c:d = 6:7$ ------(4)

make $c$ part equal Multiply (3) by 2 and (4) by (5)

$a:b:c = 16:24:30$ and $c:d = 30 : 35$

Hence $a:b:c:d = 16:24:30:35$

**Illustration 1.3.17**

A certain amount was divided between A and B in the ratio 4:3. If B's share was ₹ 2400, what is the total amount?

**Solution**

Let their shares be $4x$ and $3x$. Then, $3x = 2400 \Rightarrow x = 800$

$\therefore$Total amount = $7x = 5600$

**Illustration 1.3.18**

If 72 men complete a work in 24 days, how many men will complete the work in 36 days?

**Solution**

| Men | Days |
|---|---|
| 72 $(x_1)$ | 24 $(y_1)$ |
| $x$ $(x_2)$ | 36 $(y_2)$ |

Let $x$ men complete the work in 36 days. Compare *men* with *days*.

More days, less men. (Indirect variation)

If there are *more men*, it takes *less days* to complete the work. (*Indirect variation*)

$$72: x = 36: 24 \Longrightarrow 72 \times 24 = 36x \Longrightarrow x = 72 \times \frac{24}{36} \Longrightarrow 48$$

**Illustration 1.3.19**

There are 6 workers to paint a house. They typically paint the house in 8 hours. If 4 workers are not come to work today, how long will it take the remaining workers to paint the house.

**Solution**

If there are less workers, it takes more hours to complete the work. These two quantities are indirectly proportional to each other.

| Workers | Hours |
|---|---|
| 6 $(x_1)$ | 8 $(y_1)$ |
| 2 $(x_2)$ | $x$ $(y_2)$ |

$$x_1 : x_2 = y_2 : y_1 \Longrightarrow 6 : 2 = x : 8 \Longrightarrow 48 = 2x \Longrightarrow x = \frac{48}{2} \Longrightarrow x = 24$$

**Illustration 1.3.20**

If 300 men can complete a work in 16 days, how many men would do $\frac{1}{5}$ of the work in 15 days?

| **Men** | **Work** | **Days** |
|---|---|---|
| 300 $(x_1)$ | 1 $(y_1)$ | 16 $(z_1)$ |
| $x$ $(x_2)$ | 1/5 (($y_2$) | 15 $(z_2)$ |

**Solution:**

Compare *men* with *work* and *days*.
More *men* can do *more work*. (*Direct variation*)
If there are *more men*, it takes *less days* to complete the work. (*Indirect variation*)
Hence, it is a Combined Variation: $\frac{x_1}{x_2} = \frac{y_1}{y_2} \times \frac{z_1}{z_2}$

$$\frac{300}{x} = \frac{1}{\frac{1}{5}} \times \frac{15}{16} \Longrightarrow \frac{300}{x} = \frac{75}{16} \Longrightarrow x = 64 \text{ men}$$

**Illustration 1.3.21**

If 75 men can dig a trench 300 m long, 9 m wide and 3 m deep in 6 days working 5 hours a day. In how many days will 50 men dig another trench 540 m long,6m wide and $4\frac{1}{2}$ m deep working 4 hours a day?

**Solution:**

Let x be the number of days

| Men | Length(m) | Width(m) | Depth(m) | Time(hours) | Days |
|---|---|---|---|---|---|
| 75 | 300 | 9 | 3 | 5 | 6 |
| 50 | 540 | 6 | 4.5 | 4 | x |

Less men, more days (Indirect proportion) 50:75

More length, more days (Direct proportion) 300:540

Less width, less days (Direct proportion) 9:6

More depth, more days (Direct proportion) 3:4.5
Less time, more days (Indirect proportion) 4:5. these ratios proportional to $6 : x$

Compounding the ratios $50 \times 300 \times 9 \times 3 \times 4 \times x = 75 \times 540 \times 6 \times 4.5 \times 5 \times 6$

$$x = \frac{75 \times 540 \times 6 \times 4.5 \times 5 \times 60}{50 \times 300 \times 9 \times 3 \times 4} = 20.25\ days$$

Number of days = 20 ¼ days

**Illustration 1.3.22**

2 men and 3 boys complete a work in 16 days,5 men and 6 boys complete the same work in 7 days. In how many days 4 men and 8 boys will finish the same work.

**Solution:**

2 men and 3 boys complete the work in 16 days
i.e., 2×16 men and 3×16 boys complete the work in 1 day
32 men and 48 boys can do the work in 1 day ----------(1)
5 men and 6 boys complete the work in 7 days
i.e., 5×7 men and 6×7 boys complete the work in 1 day
35 men and 42 boys can do the work in 1 day ---------- (2)
We know (1) = (2)
i.e., 32 men + 48 boys = 35 men + 42 boys
35 men - 32 men = 48 boys - 42 boys
3men = 6 boys
i.e., 1 man = 2 boys

| Boys | Days |
|---|---|
| 7 | 16 |
| 16 | $x$ |

More boys, less days (Indirect proportion) $16{:}\,7 = 16{:}\,x \Longrightarrow 16 \times x = 7 \times 16$

$$\Longrightarrow x = 7 \times \frac{16}{16} = 7$$

Number of days = 7

**Illustration 1.3.23**

20 men complete one- fourth of a piece of work in 10 days. How many more men should be employed to finish the remaining work in 15 more days?

**Solution:**

work done = $\frac{1}{4}$; work to be done = 1- $\frac{1}{4}$ = $\frac{3}{4}$

Let the required number of men = *x*

| Work | Days | Men |
|---|---|---|
| $\frac{1}{4}$ | 10 | 20 |
| $\frac{3}{4}$ | 15 | *x* |

Compare *men* with *work* and *days*.
*Men* and *work* are directly proportional.
*Men* and *days* are indirectly proportional.

It is a combined variation. i.e.; $\frac{20}{x} = \frac{\frac{1}{4}}{\frac{3}{4}} \times \frac{15}{10} \Longrightarrow x = 40$

$$\frac{20}{x} = \frac{\frac{1}{4}}{\frac{3}{4}} \times \frac{15}{10} \Longrightarrow x = 40$$

Total 40 men. Given that, 20 men are already employed, hence 20 more men are required.

**Illustration 1.3.24**

If 5 examiners can examine a certain number of answer books in 6 days by working 4 hours a day; for how many hours a day would 6 examiners have to work in order to examine thrice the number of answer books in 10 days?

**Solution:**

Compare *hours* with *Books*, *Days* and *Examiners*. If there are *more Books*, it takes *more hours* to correct them (Direct *variation*).
If there are *more Examiners*, it takes *less hours* for them to complete the correction. (Indirect *variation)*

If there are *more Days*, *less hours* per day are required to complete the work. (*Indirect variation*)

$$\frac{4}{x} = \left(\frac{1}{3}\right) \times \left(\frac{10}{6}\right) \times \left(\frac{6}{5}\right) \Rightarrow \frac{4}{x} = \frac{2}{3} \Rightarrow x = \frac{12}{2} \Rightarrow x = 6$$

**Illustration 1.3.25**

If 8 men working 9 hours per day can complete a work in 32 days. Then 12 men working 8 hours per day, require how many days to complete the work?

**Solution**

| **Men** | **Hours** | **Days** |
|---|---|---|
| 8 | 9 | 32 |
| 12 | 8 | $x$ |

*Days* and *Men* (Indirectly *proportional*)

*Days* and *Hours* (*Indirectly proportional*)

It is an Indirect variation. i.e.; $\frac{32}{x} = \frac{12}{8} \times \frac{8}{9} \Rightarrow x = 24$ days

**Illustration 1.3.26**

25 workers construct 25 houses in 25 days. Find one worker construct one house in how many days?

**Solution**

| **Workers** | **Houses** | **Days** |
|---|---|---|
| 25 | 25 | 25 |
| 1 | 1 | $x$ |

*Days* and *Workers* (Indirect variation)
*Days* and *Houses* (Direct variation)

It is a combined variation. Because one quantity is directly proportional and another quantity is indirectly proportional to the required quantity.

i.e.; $\frac{25}{x} = \frac{1}{25} \times \frac{25}{1} \Rightarrow x = 25$

## Illustration 1.3.27

16 plumbers working 6 hours a day will finish 1600 pipes work in 18 days. 12 plumbers working 5 hours will complete 1800 pipes work in how many days?

**Solution**

Let x be the number of days

| **Plumbers** | **Hours** | **Pipes** | **Days** |
|---|---|---|---|
| 16 | 6 | 1600 | 18 |
| 12 | 5 | 1800 | $x$ |

*Days* and *Plumbers* (*Indirect proportion*)
*Days* and *Hours* (*Indirect proportion*)
*Days* and *Pipes* (*Direct proportion*)

It is combined variation.

i.e.; $\frac{18}{x} = \frac{12}{16} \times \frac{5}{6} \times \frac{1600}{1800} \Longrightarrow x = 32(\frac{2}{5})$ days

## Illustration 1.3.28

If 4 boys or 6 girls can complete a work in 11 days, in how many days same work will be completed by 6 boys and 2 girls?

**Solution:**

Let x be the number of days

Given work done by 4 boys = work done by 6 girls

i.e., 1 boy = $\frac{6}{4}$ girls = $\frac{3}{2}$ girls

6 boys and 2 girls = $6 \times \frac{3}{2} + 2 = 9 + 2 = 11$ girls

| **Girls** | **Days** |
|---|---|
| 6 | 11 |
| 11 | $x$ |

More girls, less days i.e., indirect variation

$6:11 = x:11 \Longrightarrow 6*11 = 11x \Longrightarrow x = \frac{66}{11} = 6$ days

**Illustration 1.3.29**

12 men work 10 hours a day and dig a trench in 36 days. For how many hours daily should 36 men do the work to dig it in 30 days.

**Solution:**

Let x be the hours a day men should work.

| Men | Days | Hours |
|---|---|---|
| 12 | 36 | 10 |
| 36 | 30 | $x$ |

More men less hours 36:12
Less days more hours 30:36
Compounding the ratio, we get $36 \times 30 \times x = 12 \times 36 \times 10$

$x = 12 \times 36 \times \frac{10}{36} \times 30 = 4$

Hence the men should work for 4 hours a day

## Recap

- Ratio - a way of comparing two or more quantities or values.
- Simple Ratio - A comparison of two quantities without specifying a particular unit of measurement
- Compound Ratio - Combines two or more simple ratios into one ratio
- Continued Ratio - A ratio that extends indefinitely
- Equivalent Ratios - Ratios that represent the same relationship between quantities
- Proportions - A mathematical expression that establishes the equality of two ratios
- Continued proportion - If the means of a proportion are equal
- Mean proportion- If $a: x = x: b$ then $x$ is called Mean proportion
- Direct proposition - The increase or decrease of quantities in the same ratio
- Indirect proportion - Increase in one quantity follows decrease in the other quantity in same ratio

## Objective Questions

1. Which of the two ratios $\frac{3}{4}$ and $\frac{5}{6}$ is greater?
2. Write the ratios in ascending order 3:4, 7:8. 5:6, 17:19
3. Find the compound ratio of 2:3,4:9,10:13
4. Two numbers are in the ratio of 3:4. If 5 is subtracted from each, the resulting numbers are in the ratio 2:3. Find the numbers.
5. Find value of x if 0.75: x: 5:8
6. What number should be added to each 6,14,18 and 38 so that the resulting numbers make a proportion?
7. Find which of the following ratio is the greatest?
   5:9, 13:17, 21:25, 3:7, 1:5
8. Find mean proportion of 32 and 2

## Answers

1. $\frac{5}{6}$
2. 3:4, 5:6, 7:8, 17:19
3. 80:351
4. 15 and 20
5. 1.2
6. 2
7. 21:25
8. 8

## Self-Assessment Questions

1. Discuss the importance of proportions in solving problems related to percentages and rates. Provide examples to illustrate their application.
2. Ratio between sum and difference of two numbers is 5:1. Find ratio between the numbers.

3. The ratio of the number of men to the number of women in a gathering is 5:7. If there are 96 more women than men, find the total number of people in the gathering.
4. In a mixture, the ratio of milk to water is 3:2. If 6 litters of water are added, the ratio becomes 3:4. Find the total quantity of the mixture.

## Assignments

1. Two numbers are in the ratio of 3:4. If 5 is subtracted from each, the resulting numbers are in the ratio 2:3. Find the numbers
   **Ans: 15,20**

2. If A: B = 2:3, B:C = 4:5, C:D= 6:7, then what is A: B:C:D?
   **Ans: 2:3:5:7**

3. What must be added to the numbers 10,20,30, and 50 so that the sums are proportional?

   **Ans: 10**

4. If 5 examiners can examine a certain number of answer books in 6 days by working 4 hours a day; for how many hours a day would 6 examiners have to work in order to examine thrice the number of answer books in 10 days?
   **Ans: 6**

5. If 8 men working 9 hours per day can complete a work in 32 days. Then 12 men working 8 hours per day, how many days require to complete the work?

   **Ans:24**

6. Milk and water in the ratio 5:7. Had it contained 2 litres more of milk and 2 litres less of water, the ratio would have become 4:5. What is the quantity of milk in the vessel?
   **Ans: 30**

7. If 4 men or 6 boys can finish a work in 20 days. How long will 6 men and 11 boys take to finish the same work.
   **Ans: 8**
8. 39 persons can repair a road in 12 days, working 5 hours a day. In how many days will 30 persons, working 6 hours a day, complete the work.
   **Ans: 13**

9. 16 plumbers working 6 hours a day will finish 1600 pipes work in 18 days. 12 plumbers working 5 hours will complete 1800 pipes work in how many days?
   **Ans: 32.4**

## Suggested Readings


1. Chou, Ya-Lun. *Statistical Analysis*, Holt, Rinehart and Winston, New York.
2. Croxton and Cowden. *Applied General Statistics*, Prentice Hall, London and Prentice Hall of India.
3. Croxton and Cowden. *Practical Business Statistics*, Prentice Hall, London
4. Dr. P. R Vittal – *Business Maths & Statistics*, Margham Publications

2

BLOCK

# Matrices

# Unit - 1
# Introduction

## Learning Outcomes

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

- ✓ understand the concept of Matrix.
- ✓ distinguish and comprehend various types of matrices.

## Prerequisites

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. To grasp matrices, you should have a solid foundation in several mathematical concepts. 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.

Matrices become integral when tackling systems of linear equations. Consider a scenario where optimal resource allocation must adhere to specific constraints. This scenario can be translated into a system of linear equations, and matrices emerge as a potent tool to efficiently solve such complex problems.

Suppose we have the following data about the marks of three students for three subjects

| Student | Maths | English | Science |
|---|---|---|---|
| A | 30 | 32 | 35 |
| B | 22 | 24 | 27 |
| C | 40 | 42 | 45 |

We can represent this data in matrix form as follows: $\begin{bmatrix} 30 & 32 & 35 \\ 22 & 24 & 27 \\ 40 & 42 & 45 \end{bmatrix}$

In this matrix, every row corresponds to a student (A, B, C), and each column corresponds to marks of subjects (Maths, English, Science). The entry at the $i^{th}$ row and $j^{th}$ column signifies the mark of the $i^{th}$ individual in the $j^{th}$ subject. For instance, the entry in the first row and second column (32) corresponds to the mark of student A in English. Similarly, the entry in the third row and third column (45) corresponds to the mark of student C in Science.

This matrix offers a structured method to arrange and analyse the data concerning the marks of students across different subjects. The numbers within the matrix do not inherently relate to one another. However, it empowers us to conduct operations such as determining the average mark for a specific subject or making comparisons between the marks achieved by different students.

Matrices come into play when solving systems of linear equations. For example, think about a situation where you need to allocate resources optimally given certain constraints. This can be represented as a system of linear equations, and matrices provide a powerful tool to solve such problems efficiently.


## Keywords

Square Matrix, Diagonal Matrix, Triangular Matrix, Symmetric and Skew Symmetric Matrix, Equality of Matrices


## Discussion

### 2.1.1 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 $\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix}$

Matrices are represented using capital letters such as $A, B, C$, and so on. The individual elements of a matrix are denoted as $a_{ij}$, where $i$ represents the row and $j$ represents the column. These elements can also be referred to as the $(i, j)^{th}$ elements of the matrix.

## 2.1.2 Order of a Matrix

The order of a matrix, often referred to as its size or dimensions, is a fundamental characteristic that describes how many rows and columns it contains. In essence, it provides a clear framework to understand the structure and layout of the data or information represented by the matrix.

A matrix is typically denoted by the notation $m \times n$, where "$m$" indicates the number of rows, and "$n$" indicates the number of columns. For instance, a matrix with 3 rows and 4 columns would be referred to as a "3 × 4" matrix. This ordering convention is essential as it defines the arrangement of elements within the matrix.

The order of a matrix holds significance in various mathematical and practical contexts. In linear algebra, the order of matrices determines whether they can be multiplied together.

Furthermore, in fields such as computer science and data analysis, matrices are widely used to represent data sets and transformations. The order of a matrix directly influences the operations that can be performed on it, the compatibility of different matrices for operations like addition and multiplication, and the interpretation of results in various applications.

## 2.1.3 Types of Matrices

In this section we shall discuss different types of matrices.

**Column Matrix**

A column matrix, also known as a column vector, is a type of matrix that consists of a single column of elements. In other words, it is a matrix with only one column, and its elements are arranged vertically.

The dimensions of a column matrix having $n$ elements is "$n \times 1$" where "$n$" is the number of rows and "1" is the number of columns.

For example, $\begin{bmatrix} 1 \\ 1/2 \\ 5 \end{bmatrix}$ is a column matrix of order $3 \times 1$.

$\begin{bmatrix} 6 \\ 2 \end{bmatrix}$ is a column matrix of order $2 \times 1$.

**Row Matrix**

A row matrix, also known as a row vector, is a type of matrix that has only one row but multiple columns. It is a linear arrangement of elements in a single row. Mathematically, if a row matrix has "$n$" columns, it is often represented as a $1 \times n$ matrix.

$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$ is a row matrix of order $1 \times 3$.

$\begin{bmatrix} 1 & \frac{3}{4} \end{bmatrix}$ is a row matrix of order $1 \times 2$.

**Square Matrix**

A square matrix is a type of matrix where the number of rows is equal to the number of columns. In other words, it has an equal number of rows and columns, forming a square shape. Square matrices hold a special significance in various areas of mathematics, including linear algebra and other fields.

For example, $\begin{bmatrix} 3 & 6 & 1 \\ 3 & 4 & 7 \\ 9 & 2 & 4 \end{bmatrix}$ is a square matrix of order $3 \times 3$.

$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is a square matrix of order $2 \times 2$.

**Rectangular Matrix**

A rectangular matrix is a type of matrix that does not have an equal number of rows and columns. In other words, its dimensions are such that the number of rows is not equal to the number of columns. This creates a non-square matrix, where one dimension is larger than the other.

For instance, a matrix with 3 rows and 4 columns or a matrix with 5 rows and 2 columns are both examples of rectangular matrices. In these cases, the number of rows and columns are not the same, hence the term "rectangular" to describe their shape.

For example, $\begin{bmatrix} 2 & 6 & 8 \\ 1 & 4 & 7 \end{bmatrix}$ is a rectangular matrix of order $2 \times 3$.

$$\begin{bmatrix} 3 & 6 \\ 2 & 7 \\ 8 & 9 \end{bmatrix}$$ is a rectangular matrix of order 3× 2.

**Diagonal Matrix**

A diagonal matrix is a special type of square matrix in which all the elements outside the main diagonal are zero. The main diagonal of a matrix consists of the elements that run from the top-left corner to the bottom-right corner of the matrix. In a diagonal matrix, all elements that are not on the main diagonal are explicitly set to zero. The main diagonal is called the leading diagonal.

For example, $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}$ is a diagonal matrix of order $3 \times 3$. The leading diagonal is 3,1,4.

$\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}$ is a diagonal matrix of order $2 \times 2$. The leading diagonal is $a, d$.

**Scalar Matrix**

A diagonal matrix is said to be a scalar matrix if all its diagonal entries are equal.

For example, $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ is a scalar matrix of order $3 \times 3$.

$\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$ is a scalar matrix of order $2 \times 2$.

**Zero Matrix (Null Matrix)**

A matrix is said to be a zero matrix or null matrix if each of its elements is zero.

For example, $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ is a null matrix of order $3 \times 3$.

$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ is a null matrix of order $2 \times 2$.

**Unit Matrix (Identity Matrix)**

A diagonal matrix is said to be a unit matrix (identity matrix) if each of its diagonal elements is unity.

For example,` $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is a unit matrix of order $3 \times 3$.

$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is a unit matrix of order $2 \times 2$.

The identity matrix of order $n$ is usually denoted by $I_n$ or simply by $I$.

Thus $I_1 = [1]$, $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$,

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$ are unit matrices of order 1, 2, 3 respectively.

The square matrix $[a_{ij}]$ is an identity matrix if $a_{ij} = \begin{cases} 1 \text{ when } i = j \\ 0 \text{ when } i \neq j \end{cases}$

**Triangular Matrices**

A square matrix A = $[a_{ij}]$ is said to be upper triangular matrix if $a_{ij} = 0$ for $i > j$

For example $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$ is an upper triangular matrix

A square matrix A = $[a_{ij}]$ is said to be a lower triangular matrix if $a_{ij} = 0$ for $i < j$

For example $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 5 & 6 \end{bmatrix}$ is a lower triangular matrix

**Transpose of a Matrix**

The transpose of a matrix $A$ is the matrix obtained by interchanging the rows and columns of $A$, and is denoted by $A^T$ or $A'$.

If $A$ is an $m \times n$ matrix, its transpose $A^T$ will be an $n \times m$ matrix and $(i, j)^{th}$ element in $A$ stands in the $(j, i)^{th}$ element in $A^T$.

For example, $A = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 3 & 5 \end{bmatrix}$ $then\ A^T = \begin{bmatrix} 2 & 1 \\ 4 & 3 \\ 6 & 5 \end{bmatrix}$

Let $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$, $then\ A^T = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$

**Symmetric Matrix and skew symmetric Matrix**

A square matrix $A = [aij]$ is called a symmetric matrix if $a_{ij} = a_{ji}\ for\ all\ i, j$.
Square matrix A is a symmetric matrix if $A^T = A$

For example, matrix A = $\begin{bmatrix} 3 & -1 & 1 \\ -1 & 2 & 5 \\ 1 & 5 & -2 \end{bmatrix}$ is symmetric

since $A^T = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 2 & 5 \\ 1 & 5 & -2 \end{bmatrix} = A$

A square matrix $A = [a_{ij}]$ is a skew symmetric matrix if $a_{ij} = -a_{ji}$ for all $i, j$.
A square matrix A is a skew matrix if $A^T = -A$

For example, A = $\begin{bmatrix} 0 & 2 & -3 \\ -2 & 0 & 5 \\ 3 & -5 & 0 \end{bmatrix}$ is skew symmetric

since $A^T = \begin{bmatrix} 0 & -2 & 3 \\ 2 & 0 & -5 \\ -3 & 5 & 0 \end{bmatrix} = -A$

Let A be a square matrix

(i) $A + A^T$ is a symmetric matrix

(ii) $A - A^T$ is a skew symmetric matrix

**Orthogonal Matrix**

A square matrix $A$ is called an orthogonal matrix if $A A^T = A^T A = I$

**Trace of a Matrix**

Let $A$ be the square matrix. Then the sum of all diagonal elements of $A$ is called the trace of A. Trace of matrix = $\sum_{i=1}^{n} a_{ij}$

Example: If A = $\begin{bmatrix} 1 & -5 & 7 \\ 0 & 7 & 9 \\ 11 & 8 & 9 \end{bmatrix}$ then trace of matrix = 1 + 7 + 9 =17

**Idempotent Matrix**

A square matrix A is called an idempotent matrix if $A^2 = A$

**Nil potent Matrix**

A square matrix A is called nil potent of index 2 if $A^2 = 0$

**Equality of Matrices**

Two matrices are considered equal if they have the same dimensions (i.e., the same number of rows and columns) and if their corresponding elements are equal. In other words, matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ are said to be equal if and only if they possess the same dimensions and have identical elements in the corresponding positions within the array.

For example, $\begin{bmatrix} 5 & 2 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 3 & 1 \end{bmatrix} \neq \begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix}$

If $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$, then $x = 6, y = 2$.

## Recap

- A matrix is a rectangular array of numbers arranged in rows and columns.
- Matrix consists of a single column of elements is column matrix and single row of elements is row matrix.
- A matrix in which the number of rows is equal to the number of columns - a square matrix. A matrix that does not possess an equal number of rows and columns is known as a rectangular matrix.
- A diagonal matrix is a special type of square matrix in which all the elements outside the leading diagonal are zero. If the leading diagonals are same, it is a scalar matrix. If the diagonal elements are unity it is unit matrix
- If all elements of a matrix are zero it is a zero matrix.
- Square matrix A is a symmetric matrix if $A^T = A$ and skew symmetric matrix if $A^T = -A$
- Sum of all diagonal elements of A is called the trace of A.
- A square matrix A is called an orthogonal matrix if $A A^T = A^T A = I$.
- A triangular matrix is a type of square matrix where all the elements either above or below the main diagonal are zeros. There are two main types of triangular matrices: upper triangular matrix and lower triangular matrix.
- A square matrix A is called an idempotent matrix if $A^2 = A$ and nil potent if $A^2 = 0$.

## Objective Questions

1. Which type of Matrix having the number of rows is equal to the number of columns?
2. What is the term for a square matrix that is equal to its own transpose?
3. What is the dimension of a matrix with 5 rows and 3 columns?
4. What type of matrix is characterized by having all diagonal elements the same?

## Answers

1. Square matrix
2. Symmetric matrix
3. 5× 3
4. Scalar matrix.

## Self-Assessment Questions

1. What is a square matrix?
2. What is the difference between symmetric matrix and skew symmetric matrix?
3. What is an orthogonal matrix?
4. What is idempotent and nil potent Matrix?
5. What is the condition for equality of matrix?

## Assignments

1. True or false: A 3x4 matrix can be multiplied by a 4x3 matrix. Explain your reasoning.
2. Consider the matrix A = [[1, 0, 3], [-2, 5, 0], [4, 3, 2]]. Find the transpose of A. Is A equal to its transpose? Explain why or why not.
3. Determine if the following matrix B is symmetric, skew-symmetric, or neither: B = [[0, 5, 3], [-5, 0, 1], [-3, -1, 0]] Show the transpose of B and any relevant calculations to support your answer.

## Suggested Readings

1. Sancheti and V.K.Kapoor -Business Mathematics, Sultan Chand & Sons
2. Wikes, F.M - Mathematics for Business, Finance and Economics. Thomson Learning
3. Dr. P.R. Vittal - Business Maths & Statistics, Margham Publication

# Unit - 2

# Matrix Operations

## Learning Outcomes

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

- ✓ understand to perform addition and subtraction of matrices of compatible dimensions.
- ✓ understanding matrix multiplication and how it affects the matrix's elements.
- ✓ understand the properties of matrix operations.

## Prerequisites

Operations on matrices are fundamental mathematical procedures used to manipulate, combine, and transform matrices, which are arrays of numbers organized in rows and columns. Before delving into matrix operations, a solid understanding of matrix notation, matrix arithmetic (addition, subtraction, and scalar multiplication), and matrix dimensions is crucial. These establish the groundwork for comprehending more complex matrix operations. Matrix operations include addition and subtraction, where matrices of the same dimensions are combined element-wise, following simple algebraic rules. Matrix multiplication, on the other hand, demands an understanding of the compatibility of dimensions and involves the combination of rows and columns to produce a new matrix. Further, the concept of the identity matrix- a square matrix with ones on the main diagonal and zeros elsewhere- is pivotal for matrix multiplication and serves as an analog to the number "1" in scalar arithmetic. Additionally, matrix transposition, achieved by swapping rows and columns, is an operation that finds applications in various mathematical contexts, from solving systems of linear equations to defining symmetric matrices. An understanding of matrix operations facilitates various applications, such as solving linear systems, transformations in computer graphics, data manipulation in statistics, and solving differential equations in physics and engineering.

## Keywords

Addition of Matrices, Difference of Matrices, Multiplication of Matrices

## Discussion

Matrix operations are fundamental mathematical concepts that find practical application in various fields. A matrix is essentially an organized collection of numbers, and operations performed on matrices allow us to manipulate and analyze data efficiently. Imagine you are running a retail business with multiple product lines and several branches. Matrix operations can help streamline various aspects of your business operations. One essential matrix operation is addition. In your business, you have sales data for different products across various branches. By adding matrices representing sales figures, you can quickly determine the total sales for each product category. For instance, if you have a matrix representing sales of electronic items in one branch and another matrix for sales of clothing in another branch, adding these matrices element-wise would provide a consolidated view of your total sales across all branches for different product categories.

Matrix multiplication is another powerful operation. Suppose you have matrices representing the unit cost of products and the quantities sold in each branch. Matrix multiplication can help you calculate the total cost incurred and the revenue generated by each branch. This information is vital for analyzing profitability and making informed decisions on resource allocation.

In the realm of inventory management, matrix operations play a role as well. Let's say you have a matrix representing the stock levels of different products in your various branches. By subtracting a matrix representing customer orders, you can efficiently update your inventory levels, ensuring timely restocking and preventing stockouts.

Furthermore, matrix operations can aid in decision-making. If you have matrices representing different investment options and their potential returns, matrix multiplication can help you assess the overall return on investment for each option, aiding in making sound financial choices.

In commerce, data often comes in complex formats, and matrix operations provide a systematic way to process and analyze this data. By leveraging matrix addition, multiplication, and other operations, businesses can gain insights into sales trends, optimize resource allocation, manage inventory effectively, and ultimately make informed decisions that contribute to growth and success. Thus, understanding and applying matrix operations in real-life commerce scenarios is a valuable skill for professionals in the business world.

### 2.2.1 Addition of matrices

Matrix addition is a fundamental mathematical operation that involves combining two matrices of the same dimensions by adding their corresponding elements. This operation is analogous to element-wise addition in regular arithmetic, where each element in the resulting matrix is the sum of the corresponding elements from the input matrices.

For matrix addition to be valid, the matrices must have identical numbers of rows and columns, making their sizes compatible. If the matrices have different dimensions, they are not conformable for addition, and the operation is undefined. Conformability is a key prerequisite for matrix addition and ensures that the operation can be executed properly without any ambiguity.

i.e., If $A = [\, a_{ij} \,]$ and $B = [\, b_{ij} \,]$are two matrices of the same order, their sum $A + B$ is defined as the matrix $[Cij\,]$where $c_{ij} = a_{ij} + b_{ij}$

Further, if $A$ and $B$ are not of the same order $A + B$ is not defined.

If $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ is a $3 \times 3$ matrix and

$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$ is another matrix of order $3 \times 3$.

Then we define $A + B = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \\ a_{31}+b_{31} & a_{32}+b_{32} & a_{33}+b_{33} \end{bmatrix}$

Example :Given $A = \begin{bmatrix} 6 & 43 \\ -1 & 7 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & -10 \\ 2 & 15 \end{bmatrix}$

Then $A + B = \begin{bmatrix} 6+3 & 43-10 \\ -1+2 & 7+15 \end{bmatrix} = \begin{bmatrix} 9 & 33 \\ 1 & 22 \end{bmatrix}$

**Properties of matrix addition**

**1. Matrix addition is commutative**

If $A$ and $B$ are two matrices of the same order, then $A + B = B + A$

**2. Matrix addition is associative**

If $A, B, C$ are any three matrices of the same order, then

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

**3. Existence of additive identity**

Corresponding to every $m \times n$ matrix $A$, there exists a zero matrix 0 of the same order such that $A + 0 = 0 + A = A$.

The zero matrix 0 is known as additive identity.

**4. Existence of additive inverse**

Let $A = [a_{ij}]_{m \times n}$ be any matrix, then we have another matrix as $-A = [-a_{ij}]_{m \times n}$ such that $A + (-A) = (-A) + A = 0$.

So $-A$ is the additive inverse of $A$ or negative of $A$ .

## 2.2.2 Difference of two matrices

If $A$ and $B$ be two matrices, then we write $A - B = A + (-B)$

Thus, the difference $A - B$ is obtained by subtracting from each element of $A$ the corresponding elements of $B$.

For Example, If $A = \begin{bmatrix} 1 & -10 \\ 5 & 43 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & 11 \\ 4 & 21 \end{bmatrix}$ then

$$A - B = \begin{bmatrix} 1 & -10 \\ 5 & 43 \end{bmatrix} - \begin{bmatrix} 3 & 11 \\ 4 & 21 \end{bmatrix} = \begin{bmatrix} 1 & -10 \\ 5 & 43 \end{bmatrix} + \begin{bmatrix} -3 & -11 \\ -4 & -21 \end{bmatrix}$$

$$= \begin{bmatrix} 1-3 & -10-11 \\ 5-4 & 43-21 \end{bmatrix} = \begin{bmatrix} -2 & -22 \\ 1 & 22 \end{bmatrix}$$

## 2.2.3 Scalar multiplication

Let A be a $m \times n$ matrix and $k$ be a scalar (any number). Then the $m \times n$ matrix obtained by multiplying every element of the matrix $A$ by $k$ is called the scalar multiple of $A$ by $k$ and is denoted by $kA$.

Consider the following information regarding the number of male and female workers in three organizations $A, B$ $and$ $C$.

| | M | F |
|---|---|---|
| A | 60 | 40 |
| B | 55 | 70 |
| C | 90 | 80 |

Now it is decided to double the number of male and female workers of each organization The reversed matrix is of the form

$$\begin{array}{cc} & \text{male} \quad \text{female} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} 2\times 60 & 2\times 40 \\ 2\times 55 & 2\times 70 \\ 2\times 90 & 2\times 80 \end{bmatrix} = \begin{bmatrix} 120 & 80 \\ 110 & 140 \\ 180 & 160 \end{bmatrix} \end{array}$$

Thus if , $A = [\, a_{ij}]$ then $kA = [\, ka_{ij}]$

For $example$: $If\ A = \begin{bmatrix} 2 & 1-1 \\ 3 & 4-5 \end{bmatrix}$ then $3A = \begin{bmatrix} 3\times 2 & 3\times 1 & 3\times -1 \\ 3\times 3 & 3\times 4 & 3\times -5 \end{bmatrix}$

$$= \begin{bmatrix} 6 & 3 & -3 \\ 9 & 12 & -15 \end{bmatrix}$$

## 2.2.4 Matrix multiplication

Suppose Lakshmi and Swathi are two sisters, Lakshmi wants to buy 7 notebooks and 3 pens, while Swathi needs 5 notebooks and 4 pens. The price of notebook is Rs.20 each and that of pen is Rs.8 each. Find how much money does each need to spend?

Lakshmi intends to purchase 7 notebooks at a cost of Rs. 20 each, along with 3 pens at a price of Rs. 8 each. Her total expenditure amounts to 7×20 + 3×8 = Rs. 164. On the other hand, Swathi plans to buy 5 notebooks valued at Rs. 20 each, along with 4 pens priced at Rs. 8 each. Her total expense is 5×20 + 4×8 = Rs. 132.

Now in matrix representation, we can write the above information as follows

$$\begin{array}{c} \\ \text{Lakshmi} \\ \text{Swathi} \end{array} \begin{array}{c} \text{N} \quad \text{P} \\ \begin{bmatrix} 7 & 3 \\ 5 & 4 \end{bmatrix} \end{array} \quad \begin{array}{c} \\ \text{Notebook} \\ \text{Pen} \end{array} \begin{array}{c} \text{Price} \\ \begin{bmatrix} 20 \\ 8 \end{bmatrix} \end{array}$$

Money needed (Rs)

$$\begin{bmatrix} 7\times 20+3\times 8 \\ 5\times 20+4\times 8 \end{bmatrix} = \begin{bmatrix} 140+24 \\ 100+32 \end{bmatrix} = \begin{bmatrix} 164 \\ 132 \end{bmatrix}$$

In this example we multiply two matrices. It is noted that for multiplication of two matrices A and B, the number of columns in A should be equal to the number of rows in B. For multiplication of two matrices A and B, we take rows of A and columns of B, multiply them element-wise and take the sum.

$$A = [a_{11} \quad a_{12} \quad \cdots \quad a_{1m}]\ and\ B = \begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{m1} \end{bmatrix}$$

Number of columns of $A = m$, Number of rows of $B = m$

Number of columns of $A$ = Number of rows of $B$. Hence matrices are conformable for multiplication.

AB = $[a_{11}b_{11} + a_{12}b_{21} + \cdot \cdot \cdot + a_{1m}b_{m1}]$

Hence order of $A = 1 \times m$, order of $B = m \times 1$

Order of AB = 1 × 1

These ideas can be generalized in the following definition of matrix multiplication.

**Definition:** Let $A$ be a $m \times n$ matrix and $B$ be a $n \times p$ matrix. So that number of columns of $A$ is equal to the number of rows of $B$. Then the product $C = AB$ is an $m \times p$ matrix where each element $C_{ij}$ of $C$ is obtained by multiplying corresponding elements of the i$^{th}$ row of $A$ by those of the j$^{th}$ column of $B$ and then adding the product.

$$AB = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1j} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2j} & \cdots & b_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nj} & \cdots & b_{np} \end{bmatrix}$$

$$= \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ c_{i1} & c_{i2} & \cdots & c_{ip} \\ \vdots & \vdots & \vdots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{bmatrix} = C$$

where $C_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots a_{in}b_{nj} = \sum_{k=1}^{n} a_{a_{ik}}^{b_{kj}}$

**Note:** It should carefully be noted that the product $AB$ is defined if and only if the number of columns of $A$ equals that of the number of rows of $B$.

**Properties of Matrix multiplication**

**1. Matrix multiplication is not necessarily commutative**

i.e., if $A$ and $B$ are two matrices then $AB \neq BA$

**Note:** This does not mean that $AB \neq BA$ for every pair of matrices $A, B$ for which $AB$ and $BA$, are defined.

For instance,

$$If\ A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix},\ B = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}$$

$$AB = \begin{bmatrix} 3\times 4+0\times 0 & 3\times 0+0\times 3 \\ 0\times 4+3\times 0 & 0\times 0+3\times 3 \end{bmatrix} = \begin{bmatrix} 12 & 0 \\ 0 & 9 \end{bmatrix}$$

$$BA = \begin{bmatrix} 4\times 3+0\times 0 & 4\times 0+0\times 3 \\ 0\times 3+3\times 0 & 0\times 0+3\times 3 \end{bmatrix} = \begin{bmatrix} 12 & 0 \\ 0 & 9 \end{bmatrix}$$

$$AB = BA$$

i.e., multiplication of diagonal matrices of same order will be commutative.

**2. Associative Law**

For any three matrices $A, B$ $and$ $C$ we have $(AB)C = A(BC)$

**3. Distributive Law**

For three matrices $A, B$ $and$ $C$

$$(i) A(B + C) = AB + AC$$
$$(ii) (A + B)C = AC + BC$$

**4. Existence of multiplicative identity**

For every square matrix A, there exists an identity matrix of same order such that

$$IA = AI = A$$

**Illustration 2.2.1**

Convert the linear equations $3x - 2y + z = 4$

$x - 7y + 2z = 3$

$5x + y - 4y = 1$ into matrix form

**Solution**

The coefficient of $x$, $y$ and $z$ in these equations can be expressed in the form of a 3 × 3 matrix as follows

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

**Illustration 2.2.2**

If $A = \begin{bmatrix} 3 & 2 & 5 \\ 7 & -4 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 \\ 2 & -1 \\ 3 & 5 \end{bmatrix}$ Find $AB$.

**Solution**

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

A is a $2 \times 3$ matrix, B is a $3 \times 2$ matrix

Columns of A= Rows of B

Hence A and B are conformable for multiplication

$$A \times B = \begin{bmatrix} 3 \times 1 + 2 \times 2 + 5 \times 3 & 3 \times 2 + 2 \times -1 + 5 \times 5 \\ 7 \times 1 + -4 \times 2 + 0 \times 3 & 7 \times 2 + -4 \times -1 + 0 \times 5 \end{bmatrix}$$

$$= \begin{bmatrix} 3 + 4 + 15 & 6 + -2 + 25 \\ 7 + -8 + 0 & 14 + 4 + 0 \end{bmatrix}$$

$$= \begin{bmatrix} 22 & 29 \\ -1 & 18 \end{bmatrix}$$

**Illustration 2.2.3**

If $A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$

Verify that

i. $(A')' = A$ where $A'$ is the transpose of $A$

ii. $(kA)' = kA'$

iii. $(A + B)' = A' + B'$

**Solution**

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

$A' = \begin{bmatrix} 1 & -1 \\ 2 & -2 \\ 3 & -3 \end{bmatrix}$ $\quad (A')' = \begin{bmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \end{bmatrix}$

Hence $(A')' = A$

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

$kA = \begin{bmatrix} k & 2k & 3k \\ -k & -2k & -3k \end{bmatrix}$

$(kA)' = \begin{bmatrix} k & -k \\ 2k & -2k \\ 3k & -3k \end{bmatrix}$

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

$$kA' = k \times \begin{bmatrix} 1 & -1 \\ 2 & -2 \\ 3 & -3 \end{bmatrix} = \begin{bmatrix} k & -k \\ 2k & -2k \\ 3k & -3k \end{bmatrix}$$

$\therefore \ (kA)' = kA'$

iv. $A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \end{bmatrix}, B = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$

$$A + B = \begin{bmatrix} 1+1 & 2+3 & 3+5 \\ -1+2 & -2+4 & -3+6 \end{bmatrix} = \begin{bmatrix} 2 & 5 & 8 \\ 1 & 2 & 3 \end{bmatrix}$$

$$(A+B)' = \begin{bmatrix} 2 & 1 \\ 5 & 2 \\ 8 & 3 \end{bmatrix}$$

$$A' = \begin{bmatrix} 1 & -1 \\ 2 & -2 \\ 3 & -3 \end{bmatrix} \quad B' = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

$$A' + B' = \begin{bmatrix} 1+1 & -1+2 \\ 2+3 & -2+4 \\ 3+5 & -3+6 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 5 & 2 \\ 8 & 3 \end{bmatrix}$$

$(A+B)' = A' + B'$

**Illustration 2.2.4**

$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$ Verify that $(AB)' = B'A'$

**Solution**

$$AB = \begin{bmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \times 1 + 2 \times 3 + 3 \times 5 & 1 \times 2 + 2 \times 4 + 3 \times 6 \\ -1 \times 1 \pm 2 \times 3 \pm 3 \times 5 & -1 \times 2 \pm 2 \times 4 \pm 3 \times 6 \end{bmatrix}$$

$$= \begin{bmatrix} 1 + 6 + 15 & 2 + 8 + 18 \\ -1 + -6 + -15 & -2 + -8 + -18 \end{bmatrix}$$

$$= \begin{bmatrix} 22 & 28 \\ -22 & -28 \end{bmatrix}$$

$$(AB)' = \begin{bmatrix} 22 & -22 \\ 28 & -28 \end{bmatrix}$$

$$A' = \begin{bmatrix} 1 & -1 \\ 2 & -2 \\ 3 & -3 \end{bmatrix}, B' = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

$B'A' =$

$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 2 & -2 \\ 3 & -3 \end{bmatrix} = \begin{bmatrix} 1 \times 1 + 3 \times 2 + 5 \times 3 & 1 \times -1 + 3 \times -2 + 5 \times -3 \\ 2 \times 1 + 4 \times 2 + 6 \times 3 & 2 \times -1 + 4 \times -2 + 6 \times -3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 + 6 + 15 & -1 + -6 + -15 \\ 2 + 8 + 18 & -2 + -8 + -18 \end{bmatrix} = \begin{bmatrix} 22 & -22 \\ 28 & -28 \end{bmatrix}$$

$(AB)' = B'A'$

**Illustration 2.2.5**

If $A = \begin{bmatrix} 2 & 4 & 7 \\ 3 & 1 & -7 \\ 5 & -6 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -1 & 7 \\ 8 & 2 & 5 \\ 2 & -1 & 3 \end{bmatrix}$. Find A+B.

**Solution**

$$A+B = \begin{bmatrix} 2 & 4 & 7 \\ 3 & 1 & -7 \\ 5 & -6 & 5 \end{bmatrix} + \begin{bmatrix} 1 & -1 & 7 \\ 8 & 2 & 5 \\ 2 & -1 & 3 \end{bmatrix} = \begin{bmatrix} 2+1 & 4-1 & 7+7 \\ 3+8 & 1+2 & -7+5 \\ 5+2 & -6-1 & 5+3 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 3 & 14 \\ 11 & 3 & -2 \\ 7 & -7 & 8 \end{bmatrix}$$

**Illustration 2.2.6**

If $A = \begin{bmatrix} 1 & 6 & 8 \\ 2 & 3 & -6 \\ -2 & 6 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & -4 & 6 \\ 3 & 2 & 9 \\ 6 & -2 & -3 \end{bmatrix}$. Find A-B.

**Solution**

$$A\text{-}B = \begin{bmatrix} 1 & 6 & 8 \\ 2 & 3 & -6 \\ -2 & 6 & 3 \end{bmatrix} - \begin{bmatrix} 2 & -4 & 6 \\ 3 & 2 & 9 \\ 6 & -2 & -3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 6 & 8 \\ 2 & 3 & -6 \\ -2 & 6 & 3 \end{bmatrix} + \begin{bmatrix} -2 & 4 & -6 \\ -3 & -2 & -9 \\ -6 & 2 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} -1 & 10 & 2 \\ -1 & 1 & -15 \\ -8 & 8 & 6 \end{bmatrix}$$

**Illustration 2.2.7**

If $A = \begin{bmatrix} 2 & 4 & -1 \\ 3 & 4 & 1 \\ 2 & -2 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 5 & 1 \\ 4 & -1 & 2 \\ 5 & 1 & -3 \end{bmatrix}$. Find $A \times B$.

**Solution**

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

$$= \begin{bmatrix} 2\times1+4\times4+ -1\times5 & 2\times5 + 4\times-1+ -1\times1 & 2\times1+4\times2+ -1\times -3 \\ 3\times1+4\times4+1\times5 & 3\times5+4\times-1 +1\times1 & 3\times1+4\times2+1\times-3 \\ 2\times1\pm2\times4+2\times5 & 2\times5+ -2\times-1 +2\times1 & 2\times1+ -2\times2 +2\times -3 \end{bmatrix}$$

$$= \begin{bmatrix} 13 & 5 & 13 \\ 24 & 12 & 8 \\ 4 & 14 & -8 \end{bmatrix}$$

**Illustration 2.2.8**

If $A = \begin{bmatrix} 1 & -2 & 3 \\ 2 & 3 & -1 \\ -3 & 1 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 1 & 2 & 0 \end{bmatrix}$. Find AB and BA and show that $AB \neq BA$

**Solution**

$$AB = \begin{bmatrix} 1 & -2 & 3 \\ 2 & 3 & -1 \\ -3 & 1 & 2 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 1 & 2 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1\times1-2\times0+3\times1 & 1\times0-2\times1+3\times2 & 1\times2-2\times2+3\times0 \\ 2\times1+3\times0-1\times1 & 2\times0+3\times1-1\times2 & 2\times2+3\times2-1\times0 \\ -3\times1+1\times0+2\times1 & -3\times0+1\times1+2\times2 & -3\times2+1\times2+2\times0 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & 4 & -2 \\ 1 & 1 & 10 \\ -1 & 5 & -4 \end{bmatrix}$$

$$BA = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 1 & 2 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & -2 & 3 \\ 2 & 3 & -1 \\ -3 & 1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 1\times1+0\times2+2\times-3 & 1\times-2+0\times3+2\times1 & 1\times3+0\times-1+2\times2 \\ 0\times1+1\times2+2\times-3 & 0\times-2+1\times3+2\times1 & 0\times3+1\times-1+2\times2 \\ 1\times1+2\times2+0\times-3 & 1\times-2+2\times3+0\times1 & 1\times3+2\times-1+0\times2 \end{bmatrix}$$

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

Order of AB and BA are the same but their corresponding elements are not equal. Hence AB≠BA.

**Illustration 2.2.9**

If $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -2 \\ -1 & 0 \\ 2 & -1 \end{bmatrix}$. Obtain the product AB and explain why BA is not defined.

**Solution**

$$AB = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix} \times \begin{bmatrix} 1 & -2 \\ -1 & 0 \\ 2 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 0\times1+1\times-1+2\times2 & 0\times-2+1\times0+2\times-1 \\ 1\times1+2\times-1+3\times2 & 1\times-2+2\times0+3\times-1 \\ 2\times1+3\times-1+4\times2 & 2\times-2+3\times0+4\times-1 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & -2 \\ 5 & -5 \\ 7 & -8 \end{bmatrix}$$

BA is not defined because the number of columns of B≠ number of rows of $A$.

**Illustration 2.2.10**

Prove that $\frac{1}{3}\begin{bmatrix} -2 & 1 & 2 \\ 2 & 2 & 1 \\ 1 & -2 & 2 \end{bmatrix}$ is an orthogonal matrix.

**Solution**

A matrix is orthogonal when $AA^T = A^T A = I$

$$A = \frac{1}{3}\begin{bmatrix} -2 & 1 & 2 \\ 2 & 2 & 1 \\ 1 & -2 & 2 \end{bmatrix}, \ A^T = \frac{1}{3}\begin{bmatrix} -2 & 2 & 1 \\ 1 & 2 & -2 \\ 2 & 1 & 2 \end{bmatrix}$$

$$AA^T = \frac{1}{9}\begin{bmatrix} -2 & 1 & 2 \\ 2 & 2 & 1 \\ 1 & -2 & 2 \end{bmatrix}\begin{bmatrix} -2 & 2 & 1 \\ 1 & 2 & -2 \\ 2 & 1 & 2 \end{bmatrix}$$

$$= \frac{1}{9}\begin{bmatrix} 4+1+4 & -4+2+2 & -2-2+4 \\ -4+2+2 & 4+4+1 & 2-4+2 \\ -2-2+4 & 2-4+2 & 1+4+4 \end{bmatrix}$$

$$= \frac{1}{9}\begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= I$$

$$A^T A = \frac{1}{9}\begin{bmatrix} -2 & 2 & 1 \\ 1 & 2 & -2 \\ 2 & 1 & 2 \end{bmatrix}\begin{bmatrix} -2 & 1 & 2 \\ 2 & 2 & 1 \\ 1 & -2 & 2 \end{bmatrix}$$

$$= \frac{1}{9}\begin{bmatrix} 4+4+1 & -2+4-2 & -4+2+2 \\ -2+4-2 & 1+4+4 & 2+2-4 \\ -4+2+2 & 2+2-4 & 4+1+4 \end{bmatrix}$$

$$= \frac{1}{9}\begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$= I$

Here, $A A^T = A^T A = I$. So, the matrix is orthogonal.

**Illustration 2.2.11**

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

**Solution**

Trace of the matrix $A = 4 + 3 + 2 = 9$

**Illustration 2.2.12**

Express the matrix $\begin{bmatrix} -1 & 7 & 1 \\ 2 & 3 & 4 \\ 5 & 0 & 5 \end{bmatrix}$ as the sum of a symmetric matrix and a skew symmetric matrix.

**Solution**

Given Matrix is $A = \begin{bmatrix} -1 & 7 & 1 \\ 2 & 3 & 4 \\ 5 & 0 & 5 \end{bmatrix}$ $\quad \therefore A^T = \begin{bmatrix} -1 & 2 & 5 \\ 7 & 3 & 0 \\ 1 & 4 & 5 \end{bmatrix}$

$$A + A^T = \begin{bmatrix} -1 & 7 & 1 \\ 2 & 3 & 4 \\ 5 & 0 & 5 \end{bmatrix} + \begin{bmatrix} -1 & 2 & 5 \\ 7 & 3 & 0 \\ 1 & 4 & 5 \end{bmatrix}$$

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

$$= \begin{bmatrix} -2 & 9 & 6 \\ 9 & 6 & 4 \\ 6 & 4 & 10 \end{bmatrix}$$

$B = \frac{1}{2}(A + A^T) = \begin{bmatrix} -1 & \frac{9}{2} & 3 \\ \frac{9}{2} & 3 & 2 \\ 3 & 2 & 5 \end{bmatrix}$ is a symmetric matrix

$$A - A^T = \begin{bmatrix} -1+1 & 7-2 & 1-5 \\ 2-7 & 3-3 & 4-0 \\ 5-1 & 0-4 & 5-5 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 5 & -4 \\ -5 & 0 & 4 \\ 4 & -4 & 0 \end{bmatrix}$$

$$C = \frac{1}{2}(\ A1A^T) = \begin{bmatrix} 0 & \frac{5}{2} & -2 \\ -\frac{5}{2} & 0 & 2 \\ 2 & -2 & 0 \end{bmatrix}$$ is a skew symmetric matrix

$$B + C = \begin{bmatrix} -1 & \frac{9}{2} & 3 \\ \frac{9}{2} & 3 & 2 \\ 3 & 2 & 5 \end{bmatrix} + \begin{bmatrix} 0 & \frac{5}{2} & -2 \\ -\frac{5}{2} & 0 & 2 \\ 2 & -2 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 7 & 1 \\ 2 & 3 & 4 \\ 5 & 0 & 5 \end{bmatrix} = A.$$

**Illustration 2.2.13**

Find $x, y, z$ $and$ $w$ $if$ $3\begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} x & 6 \\ -1 & 2w \end{bmatrix} + \begin{bmatrix} 4 & x+y \\ z+w & 3 \end{bmatrix}$

**Solution**

The given equation is

$$\begin{bmatrix} 3x & 3y \\ 3z & 3w \end{bmatrix} = \begin{bmatrix} x & 6 \\ -1 & 2w \end{bmatrix} + \begin{bmatrix} 4 & x+y \\ z+w & 3 \end{bmatrix}$$

$$= \begin{bmatrix} x+4 & 6+x+y \\ -1+z+w & 2w+3 \end{bmatrix}$$

Since the matrices are equal, equating the corresponding elements on the two sides

$3x = x + 4, \quad 3y = 6 + x + y, \quad 3z = -1 + z + w, \quad 3w = 2w + 3,$

$2x = 4, \quad 2y = 6 + x, \quad 2z = -1 + w, \quad 2w = w + 3,$

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

$x = 2, \quad y = 4, \quad z = 1, \quad w = 3$

**Illustration 2.2.14**

If $A = \begin{bmatrix} 3 & 2 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}$ Find $A^2 - 7A + 10$.

**Solution**

$$A^2 = \begin{bmatrix} 3 & 2 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}\begin{bmatrix} 3 & 2 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix} = \begin{bmatrix} 11 & 14 & 0 \\ 7 & 18 & 0 \\ 0 & 0 & 25 \end{bmatrix}$$

$$A^2 - 7A + 10I = \begin{bmatrix} 11 & 14 & 0 \\ 7 & 18 & 0 \\ 0 & 0 & 25 \end{bmatrix} - 7\begin{bmatrix} 3 & 2 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix} + 10\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 11 - 21 + 10 & 14 - 14 + 0 & 0 \\ 7 - 7 + 0 & 18 - 28 + 10 & 0 \\ 0 & 0 & 25 - 35 + 10 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0$$

## Recap

- Matrix addition and its properties
- Difference of two matrices
- Matrix multiplication and its properties

## Objective Questions

1. If A= $\begin{bmatrix} 4 & x+2 \\ 2x-3 & x+1 \end{bmatrix}$ is symmetric, then what is the value of $x$?

2. If $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ what is $A^4$

3. If $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $= B \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$. Find $A + B$.

4. Let $A = \begin{bmatrix} 2 & -3 \\ 1 & 4 \end{bmatrix}$ and $= \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix}$. Is it possible to form the product $AB$

5. If $\begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} p \\ 0 \end{bmatrix}$, then what is $p$?

## Answers

1. 5
2. $A^4 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
3. $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$
4. No. Number of columns of $A \neq$ Number of rows of $B$
5. $p = -1$

## Self-Assessment Questions

1. If $A = \begin{bmatrix} 1 & 5 & 1 & 3 \\ 2 & 1 & 0 & 5 \\ 7 & 1 & 8 & -7 \\ 0 & 2 & 1 & 6 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ Show that $AB = BA$

   Ans: $\begin{bmatrix} 1 & 5 & 1 & 3 \\ 2 & 1 & 0 & 5 \\ 7 & 1 & 8 & -7 \\ 0 & 2 & 1 & 6 \end{bmatrix}$

2. Express $\begin{bmatrix} 2 & 5 & -7 \\ -9 & 12 & 4 \\ 15 & -13 & 6 \end{bmatrix}$ as the sum of a lower triangular and an upper triangular matrix with zero leading diagonal.

   Ans: L= $\begin{bmatrix} 2 & 0 & 0 \\ -9 & 12 & 0 \\ 15 & -13 & 0 \end{bmatrix}$ U= $\begin{bmatrix} 0 & 5 & -7 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix}$

3. Evaluate $A^2 - 3A + 9I$, if $I$ is the unit matrix of order 3 and

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

   Ans: L= $\begin{bmatrix} -6 & 1 & 2 \\ 5 & 4 & 4 \\ 2 & 8 & -3 \end{bmatrix}$

4. Express the matrix $\begin{bmatrix} 4 & 2 & -3 \\ 1 & 3 & -6 \\ -5 & 0 & -7 \end{bmatrix}$ as the sum of a symmetric matrix and a skew symmetric matrix.

   Ans: $\begin{bmatrix} 4 & \frac{3}{2} & -4 \\ \frac{3}{2} & 3 & -3 \\ -4 & -3 & -7 \end{bmatrix} + \begin{bmatrix} 0 & \frac{1}{2} & 1 \\ -\frac{1}{2} & 0 & -3 \\ -1 & 3 & 0 \end{bmatrix}$

5. If $A, B$ and $C$ are matrices of same order, then prove that $(A + B) + C = A + (B + C)$

   Where $A = \begin{bmatrix} 3 & 2 & 1 \\ -1 & 0 & 4 \\ 5 & -3 & 8 \end{bmatrix}$, $B = \begin{bmatrix} 2 & -3 & 0 \\ 5 & -1 & 7 \\ 1 & 4 & 3 \end{bmatrix}$, $C = \begin{bmatrix} 4 & 2 & 5 \\ -2 & 1 & 3 \\ 0 & 8 & -1 \end{bmatrix}$

Ans: $\begin{bmatrix} 9 & 1 & 6 \\ 2 & 0 & 14 \\ 6 & 9 & 10 \end{bmatrix}$

6. Let A=$\begin{bmatrix} 2 & 3 & 1 \\ 0 & -1 & 5 \end{bmatrix}$ and B =$\begin{bmatrix} 1 & 2 & -6 \\ 0 & -1 & 3 \end{bmatrix}$. Find $3A - 4B$.

   Ans=$\begin{bmatrix} 2 & 1 & 27 \\ 0 & 1 & 3 \end{bmatrix}$

7. Given the matrices If $A, B$ and $C$ such that A=$\begin{bmatrix} 2 & 3 & 1 \\ 3 & 0 & 2 \end{bmatrix}$ , B = $\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$,

   $\mathrm{C} = [1 \quad -2]$. Verify $(AB)C = A(BC)$.

   $\text{Ans} = \begin{bmatrix} 3 & -6 \\ 7 & -14 \end{bmatrix}$

8. If $\begin{bmatrix} -4 & x+5 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} y+4 & 2 \\ 3 & z \end{bmatrix}$, find $x, y\ and\ z.$

   $\text{Ans}; \mathrm{x} = -3, \mathrm{y} = -8, \mathrm{z} = 3$

## Suggested Readings


1. Sancheti and V.K.Kapoor -*Business Mathematics*, Sultan Chand & Sons
2. Wikes, F.M - *Mathematics for Business*, Finance and Economics. Thomson Learning
3. Dr. P.R. Vittal - *Business Maths & Statistics*, Margham Publication

3

BLOCK

# Measures of Central Tendency

# Unit – 1
# Average

## Learning Outcomes

After completing this unit, learners will be able to:

- ✓ develop proficiency in calculating various types of averages
- ✓ identify the advantages and limitations of averages
- ✓ understand the requisites needed for averages
- ✓ recognize how averages are used in real life scenarios

## Prerequisites

The use of averages is a common statistical practice to represent the central value of a variable that experiences fluctuations over time. In situations where prices, such as the price of LPG, experience significant fluctuations, we can utilize averages to express a single figure that represents the central value. By calculating the average, we obtain a value that provides a general indication of the variable's price based on a range of prices observed throughout a given period. This approach is also applicable to other variables such as income, age, weight, height, marks, exports, imports, prices, and demand. Averaging enables us to summarize the data and gain insight into the central tendency of these variables. It provides a useful indicator that helps in making comparisons, identifying trends, and making informed decisions based on a representative value.

## Keywords

Mathematical average, Positional average

## Discussion

### 3.1.1 Measuring Central tendency

The term measure of central tendency refers to a single number that is used to identify the central position within a set of data in order to define it. As a result, measures of central tendency are also known as measures of central location or average. Average is considered the most representative value in a set of data because it is determined where the concentration of items or values are highest, indicating the highest frequency on a distribution scale. As a result, the average serves as a simple representation of a diverse range of individual values, effectively capturing the essence of a complex distribution. Averages are the values that give us an indication of the concentration of observations in the central part of the distribution. It is a single value that reflects an entire set of data, around which most of the values of the data cluster.

According to Professor Arthur Lyon Bowley, Averages are, "statistical constants which enable us to comprehend in single effort, the significance of the whole".

According to Croxton and Cowden, "An average value is a single value within the range of the data that is used to represent all the values in the series".

### 3.1.1.1 Objectives

The following are the objectives of averages:

i. To give a quick overview of the data- It aids in providing a succinct explanation of the data's principal feature.
ii. Facilitates comparison- It assists in the reduction of data to a single value for comparative investigations.
iii. Helps in decision making- The majority of firms employ central tendency measurement to plan and develop their businesses' economy.
iv. Formulation of policies- Many governments as well as business organisations rely on this medium while framing policies.
v. Base for other statistical analysis- Averages forms the base for other statistical tools including mean deviation, co-efficient of variation, correlation, time series analysis, and index numbers which we will discuss later in this unit.

### 3.1.1.2 Features

An average is the representation of the aggregate data. A good average has the following characteristics.

i. Rightly and rigidly defined- The measure's definition should be crystal clear and consistent across the board and should be capable of being interpreted.
ii. Simple to calculate and understand- It should result in the same meaning for whoever is doing the mathematics.
iii. Easy to understand- It should be simple to interpret.
iv. Based on all the observations- It means that whatever measure of central tendency is used, the information it delivers must be simply comprehended.

v. A few extreme values should not have much of an impact- The value of a good average should not be harmed by a few extremely small or very large values.
vi. It should be able to be subjected to more algebraic treatment- On the basis of the average's outcome, an analyst should be able to perform more and more mathematical analysis. The average should have a well-built mathematical model for this.

### 3.1.1.3 Importance/ Advantages of Average

The average enables us to condense everything into a single number. Averages have the following advantages:

i. It simplifies data complexity.
ii. It facilitates comparative study.
iii. It is easy to remember.
iv. It is possible to develop a mathematical connection between various variables.
v. It is an important quantitative technique in research work.
vi. It aids in the condensing of information.
vii. It gives a good impression of the data collected.
viii. It aids research through statistical inference.
ix. It helps in decision making.

### 3.1.1.4 Limitations of Average

Averages are useful in statistical and mathematical calculations, but they have several drawbacks. The following are some of the major drawbacks of averages:

i. The measurements of central tendency may not always give a unique representation of the data.
ii. An average can result in a figure that is not part of the series.
iii. It could result in a ludicrous and practically impossible result.
iv. When computing the mean, it's possible to have an unrepresentative result if there are any extreme values.
v. It does not reveal the full story of a phenomenon.
vi. The spread or variability of the data is not revealed by measures of central tendency.
vii. The measures of central tendency are inappropriate for categorical or qualitative data without a numerical value. They are mostly used for quantitative data.

### 3.1.1.5 Requisites

According to Prof. Yule, an ideal average or measure of central tendency must meet the following characteristics:

i. **It should be rigidly defined** - The definition should be simple and obvious such that it can only have one interpretation among various people. Nothing in the definition should be left up to the observer's or investigator's interpretation. If it is not strictly defined, the bias that the researcher introduces will cause its value to fluctuate and making it unrepresentative of the distribution.
ii. **It should be as unaffected by sampling fluctuations** - Average should not be heavily influenced by random variations or outliers in the data. This is important because sampling fluctuations can occur due to random chance, and it's desirable for the measure of central tendency to remain stable and representative of the underlying distribution.
iii. **It should be simple to comprehend and calculate** - It should be simple to understand, and must be able to be computed quickly and easily without the need to perform complex mathematical computations. This shouldn't be done at the expense of accuracy or any other benefits that an average person may have, though.
iv. **It should be based on all the observations**- There should be no information loss as a result of not using the available data while computing an ideal average because the whole set of data at our disposal should be utilised. The average will obviously not be reflective of the distribution if the entire set of data is not used to calculate it.
v. **It should be able to be treated mathematically in the future** -Average should possess some important and interesting mathematical properties so that its use in further statistical theory is enhanced.
vi. **It should avoid giving misleading results** - Measures of central tendency should accurately represent the data and provide meaningful information. They should not lead to misleading interpretations or misrepresent the underlying distribution.

### 3.1.1.6 Types of Average

There are three types of averages.

**i. Mathematical Average**

The term "mathematical average" refers to any average in which a figure is extracted from a particular series using mathematical procedures to represent the entire series. The mathematical average includes:

a. Arithmetic mean
b. Geometric mean
c. Harmonic mean

**ii. Positional Average**

The term "positional average" refers to an average derived from a series of observations in which single number from the series based on its position is chosen to represent the entire series.

The following are some positional averages.

a. Median
a. Mode
b. Quartiles
c. Deciles
d. Percentiles
e. Septiles
f. Octiles
g. Quintiles

**iii. Miscellaneous Average**

Miscellaneous average refers to any average that is not mathematical or positional. The following are some miscellaneous averages.

a. Moving average
b. Progressive average
c. Weighted average
d. Quadratic average

The various mathematical averages and positional averages are discussed in detail in the forthcoming units.

## Recap

- Measure of central tendency- summarises data into a single value that can be used to represent all of it.
- Mathematical average- average in which a figure is extracted from a particular series using mathematical procedures to represent the entire series.
- Positional average- average derived from a series of observations in which a single number from the series based on its position is chosen to represent the entire series.
- Miscellaneous average- average that is not mathematical or positional.

## Objective Questions

1. Which term is used to denote a single number that is used to identify the central position within a set of data in order to define it?
2. List out the most commonly used mathematical averages.
3. Identify the average in which a figure is extracted from a particular series using mathematical procedures to represent the entire series.
4. Name any one miscellaneous average.
5. What is the name of the average calculated from a set of observations in which a single number based on its position is picked to represent the whole set?
6. What is miscellaneous average?
7. Give an example of positional average.

## Answers

1. Measures of central tendency
2. Arithmetic mean, Geometric mean, Harmonic mean
3. Mathematical average
4. Moving average
5. Positional average
6. Average that is not mathematical or positional
7. Median

## Self-Assessment Questions

1. Explain the term Average.
2. What are the objectives of measures of central tendency?
3. State the importance of average.
4. Explain the various types of averages.
5. "An ideal average or measure of central tendency must meet certain characteristics"- Comment.
6. Write a short note on the features of averages.

## Assignments

1. In a classroom setting, if you want to know the typical performance of students, the measure of central tendency can be used. Explain how average proves helpful in this scenario. (Hint: Under performance and Over performance can be determined)
2. List 5 real life scenarios where the measures of central tendency can be used.

## Suggested Readings


1. Sancheti and V.K.Kapoor (2014). *Business Mathematics*. Sultan Chand & Sons
2. Wikes, F.M (1998). *Mathematics for Business, Finance and Economics.* Thomson Learning
3. Dr. P.R. Vittal (2012). *Business Maths & Statistics*. Margham Publications
4. C.R.Kothari (2013). *Quantitative Techniques*. Vikas Publishing House
5. S.P.Guptha-(2012). *Statistical Methods*. Sultan Chand &Sons,New Delhi
6. C. B Gupta & Vijay Gupta (2004). *An Introduction to Statistical Methods*. Vikas Publishing
7. Goel & Goel (2014). *Mathematics & Statistics*. Taxmann Allied Services Pvt Ltd

# Unit - 2
# Mathematical Averages

## Learning Outcomes

After completing this unit, learners will be able to:

- ✓ familiarise with the various types of mathematical average
- ✓ develop an understanding of which type of average would be the most useful in a particular situation
- ✓ solve problems of arithmetic mean, geometric mean, and harmonic mean

## Prerequisites

In the field of statistics, we often need to choose a single value that represents an entire set of data. This value is known as a "measure of central tendency" because it captures the central focus of the data. Let's consider a student's report card, which includes an aggregate percentage representing their overall performance. Why is this aggregate shown on the report card? Well, it's simply to summarise the student's total marks across different subjects in a single value. By tabulating the marks obtained in each subject, we gain insights into the student's performance.

Now, suppose a student scores 845 marks out of 1000. How can they determine if it's a good score or not? One way is to compare their marks with a representative value calculated using the overall marks secured by all the students in the entire class. This is where the concepts of mathematical average come into play. This unit explains the various types of mathematical averages.

## Keywords

Arithmetic mean, Geometric mean, Harmonic mean

## Discussion

### 3.2.1 Mathematical Average

The term "mathematical average" refers to any average in which a figure is extracted from a particular series using mathematical procedures to represent the entire series. The selection of appropriate average depends on the nature of the data involved. The mathematical averages that are discussed in detail in this unit include:

a. Arithmetic mean
b. Geometric mean
c. Harmonic mean

### 3.2.2 Arithmetic Mean

The arithmetic mean (also known as the mean) is a widely used measure of central tendency. When we talk about a data's 'average,' we usually mean its arithmetic mean. An arithmetical average may be defined as "the quotient obtained by dividing the total of the values of a variable by the total number of their observations". The arithmetic mean is effective when there is an additive relationship between the numbers, which occurs frequently in 'linear' relationships when the numbers, when plotted, either lie on or near a straight line.

**Properties of Arithmetic Mean**

The arithmetic mean of a distribution has the following mathematical properties.

i. The sum of item deviations from the Arithmetic mean in a data set is always zero.

   i.e., $\sum(x - \bar{x}) = 0$

ii. The sum of squares of the deviations of the items in a data is the least when the deviation is taken from the arithmetic mean.

   i.e., $\sum(x - a)^2$ is least when a = $\bar{x}$

iii. If the mean of n observations, $x_1, x_2, \ldots\ldots, x_n$ is $\bar{x}$, then the mean of the observation,

$$(x_1 \pm a), (x_2 \pm a), \ldots\ldots, (x_n \pm a) is (\bar{x} \pm a).$$

iv. The mean of n observation, $x_1, x_2, \ldots, x_n$ is $\bar{x}$, if each observation is multiplied by p, p≠0, then the mean of the new observation is $p\bar{x}$.

**Advantages of Arithmetic Mean**

i. It has a rigid definition.
ii. It is simple to comprehend and compute.
iii. It is based upon all the observations.

iv. It is least affected by sampling fluctuations.
v. It can be subjected to further mathematical analysis.

**Disadvantages of Arithmetic Mean**

i. Extreme value has a significant impact for calculating mean.
ii. It cannot be determined by inspection.
iii. It cannot be used to measure qualitative characteristics like honesty, beauty, cleverness, and so on.
iv. It is impossible to calculate for open-ended classes.
v. It is not suitable for averaging ratios and percentages

**Computation of Arithmetic Mean**

Usually mean is denoted by the symbol '$\bar{x}$' (read as 'x bar')

**A. For Individual Series**

There are two ways for calculating the mean for individual data series.

**i. Direct method**

$$\bar{x} = \frac{\sum x}{N}$$

where, $\sum x$ – sum of observations

$N$- Number of observations.

**ii. Short Cut method**

This method uses an assumed mean and deviations taken from that assumed mean to determine the arithmetic average. 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.

$$\bar{x} = A + \frac{\sum dx}{N}$$

where, $\bar{x}$- Arithmetic mean

A- Assumed mean

$dx$- Deviation of the observations from the assumed mean $(x - A)$

$N$- 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**

**Direct method**

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

**Short Cut method**

**Assumed Average, A = 30**

| $x$ | $dx = x - A$ |
|---|---|
| 45 | 15 |
| 40 | 10 |
| 37 | 7 |
| 18 | -12 |
| 17 | -13 |
| 35 | 5 |
| 10 | -20 |
| 28 | -2 |
| 36 | 6 |
| 47 | 17 |
| | $dx = $ **13** |

$$\bar{x} = A + \frac{\sum dx}{N}$$

$$= 30 + \frac{13}{10} = 31.3$$

**Illustration 3.2.2**

The marks obtained by 10 students in a class test are given below.

| Roll No: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Marks: | 60 | 30 | 70 | 50 | 40 | 50 | 60 | 20 | 70 | 30 |

Calculate the arithmetic mean

**Solution**

**Direct method**

$$\overline{x} = \frac{\sum x}{N} = \frac{60+30+70+50+40+50+60+20+70+30}{10}$$

$$= \frac{480}{10} = 48$$

**Short cut method**

Assumed Average = 50

| Roll No | Mark (x) | dx =(x-50) |
|---|---|---|
| 1 | 60 | 10 |
| 2 | 30 | -20 |
| 3 | 70 | 20 |
| 4 | 50 | 0 |
| 5 | 40 | -10 |
| 6 | 50 | 0 |
| 7 | 60 | 10 |
| 8 | 20 | -30 |
| 9 | 70 | 20 |
| 10 | 30 | -20 |
| **∑dx** | | **-20** |

$$\overline{x} = A + \frac{\sum dx}{N}$$

$$= 50 + \frac{-20}{10}$$

= 50-2
= 48

**B. For discrete series**

**i. Direct method**

$$\bar{x} = \frac{\sum(f \times x)}{N}$$

where, $\bar{x}$ - Arithmetic mean

$x$ – Value of the variable

$f$ – Frequency

$N$ – Total number of observations, ie., $\sum f$

**ii. Short cut method**

The formula for calculating the arithmetic mean for the short cut method is,

$$\bar{x} = A + \frac{\sum(f \times dx)}{N}$$

Where,

$A$ – Assumed mean

$dx$ – $(x - A)$

f – Frequency

N – Total frequency

**Illustration 3.2.3**

The weight of 20 diabetic patients was gathered from a primary health centre by a researcher. The weights of 20 diabetic individuals are shown in the table below.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Weight (in kg): | 49 | 53 | 54 | 55 | 66 | 68 | 70 | 80 |
| No. of patients: | 1 | 2 | 4 | 5 | 3 | 2 | 2 | 1 |

Calculate the mean weight of the diabetic patients.

**Solution**

**Direct method**

| Weight (in kg) | No of diabetic patients (f) | $f \times x$ |
|---|---|---|
| 49 | 1 | 49 |
| 53 | 2 | 106 |
| 54 | 4 | 216 |
| 55 | 5 | 275 |
| 66 | 3 | 198 |
| 68 | 2 | 136 |
| 70 | 2 | 140 |
| 80 | 1 | 80 |
| **Total** | **N= 20** | **1200** |

$\bar{x} = \frac{\Sigma(f \times x)}{N} = \frac{1200}{20} = 60$

Mean weight = 60 kg

**Short cut method**

Assumed Average = 65

| Weight (in kg) $x$ | No. of diabetic patients $f$ | $dx = x - 65$ | $f \times dx$ |
|---|---|---|---|
| 49 | 1 | -16 | -16 |
| 53 | 2 | -12 | -24 |
| 54 | 4 | -11 | -44 |
| 55 | 5 | -10 | -50 |
| 66 | 3 | 1 | 3 |
| 68 | 2 | 3 | 6 |
| 70 | 2 | 5 | 10 |
| 80 | 1 | 15 | 15 |
| | $\sum f = 20$ | | $\sum f \times dx = -100$ |

$$\bar{x} = A + \frac{\Sigma(f \times dx)}{N}$$

$$= 65 - \frac{100}{20}$$

$$= 65 - 5$$

$$= 60$$

Mean weight = 60 kg

**Illustration 3.2.4**

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**

**Direct method**

| **Height (x)** | **No of employees (f)** | $f \times x$ |
|---|---|---|
| 60 | 5 | 300 |
| 62 | 10 | 620 |
| 63 | 12 | 756 |
| 65 | 18 | 1170 |
| 67 | 15 | 1005 |
| 68 | 10 | 680 |
| **Total** | **70** | **4531** |

$$\bar{x} = \frac{\Sigma(f \times x)}{N}$$

$$= \frac{4531}{70}$$

$= 64.728$(In inches)

**Short Cut Method**

**Assumed Average = 63**

| **Height (x)** | **No of employees (f)** | **dx (x-63)** | **F dx** |
|---|---|---|---|
| 60 | 5 | -3 | -15 |
| 62 | 10 | -1 | -10 |
| 63 | 12 | 0 | 0 |
| 65 | 18 | 2 | 36 |
| 67 | 15 | 4 | 60 |
| 68 | 10 | 5 | 50 |
| **Total** | **70** | | **121** |

$\bar{x} = A + \frac{\sum fdx}{N}$

$= 63 + \frac{121}{70}$

$= 63+1.728$

= 64.728 (in inches)

## C. For Continuous Series

### i. Direct method

The formula for calculating the mean of a continuous series using the direct method is as follows:

$$\bar{x} = \frac{\Sigma(f \times x)}{N}$$

Where,

$x$ = mid value = $\frac{\text{Lower limit+Upper limit}}{2}$

$N$ = Total Frequency

### ii. Short Cut Method

The formula for calculating the mean of a continuous series using the short cut method is as follows:

$$\bar{x} = A + \frac{\Sigma(f \times dx)}{N}$$

Where,

$A$ = Assumed mean

$dx = x - A$

$f$ –Frequency

$N$ - Total frequency

**Illustration 3.2.5**

The income of 60 farmers is shown in the table. Determine what their average income is.

| Income: | 10000-20000 | 20000-30000 | 30000-40000 | 40000-50000 | 50000-60000 | 60000-70000 |
|---|---|---|---|---|---|---|
| No. of farmers: | 16 | 18 | 12 | 7 | 4 | 3 |

**Solution**

**Direct method**

| **Income** | **Mid value (x)** | **f** | **f x** |
|---|---|---|---|
| 10000-20000 | 15000 | 16 | 240000 |
| 20000-30000 | 25000 | 18 | 450000 |
| 30000-40000 | 35000 | 12 | 420000 |
| 40000-50000 | 45000 | 7 | 315000 |
| 50000-60000 | 55000 | 4 | 220000 |
| 60000-70000 | 65000 | 3 | 195000 |
| **Total** | | **60** | **1840000** |

$\overline{x} = \frac{\sum fx}{N} = \frac{1840000}{60} = 30666.67$

**Short Cut Method**

**Assumed Mean = 35000**

| **Income** | **Mid value ($x$)** | $f$ | $dx$ | $fdx$ |
|---|---|---|---|---|
| 10000-20000 | 15000 | 16 | -20000 | -320000 |
| 20000-30000 | 25000 | 18 | -10000 | -180000 |
| 30000-40000 | 35000 | 12 | 0 | 0 |
| 40000-50000 | 45000 | 7 | 10000 | 70000 |
| 50000-60000 | 55000 | 4 | 20000 | 80000 |
| 60000-70000 | 65000 | 3 | 30000 | 90000 |
| **Total** | | **60** | | **-260000** |

$$\bar{x} = A + \frac{\Sigma(f \times dx)}{N}$$

$$= 35000 - \frac{260000}{60}$$

$$= 30666.67$$

**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**

**Direct method**

| Weight | Mid Value (x) | $f$ | $fx$ |
|---|---|---|---|
| 65-84 | 74.5 | 9 | 670.5 |
| 85-104 | 94.5 | 10 | 945 |
| 105-124 | 114.5 | 17 | 1946.5 |
| 125-144 | 134.5 | 10 | 1345 |
| 145-164 | 154.5 | 5 | 772.5 |
| 165-184 | 174.5 | 4 | 698 |
| 185-204 | 194.5 | 5 | 972.5 |
| **Total** | | **60** | **7350** |

**Solution**

$$\bar{x} = \frac{\Sigma(x \times f)}{N}$$

$= \frac{7350}{60}$

$= 122.5$

**Short Cut Method**

**Assumed Mean A = 134.5**

| **Weight** | **Mid Value (x)** | **f** | $dx$ | $fdx$ |
|---|---|---|---|---|
| 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** |

$\bar{x} = A + \frac{\Sigma(f \times dx)}{N}$

$= 134.5 - \frac{720}{60}$

$= 122.5$

**Illustration 3.2.7**

The output of 55 workers chosen to test the effectiveness of a certain machine is shown in the table below.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Output: | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
| No. of Workers: | 5 | 7 | 17 | 12 | 5 | 2 | 7 |

Calculate the workers' average output.

**Solution**

**Direct method**

| **Output** | **Mid Values ($x$)** | $f$ | $fx$ |
|---|---|---|---|
| 0-10 | 5 | 5 | 25 |
| 10-20 | 15 | 7 | 105 |
| 20-30 | 25 | 17 | 425 |
| 30-40 | 35 | 12 | 420 |
| 40-50 | 45 | 5 | 225 |
| 50-60 | 55 | 2 | 110 |
| 60-70 | 65 | 7 | 455 |
| **Total** | | **55** | **1765** |

$$\bar{x} = \frac{\Sigma(fx)}{N}$$
$$= \frac{1765}{55}$$
$$= 32.09$$

**Short Cut Method**

**Assumed mean is 35**

| **Output** | **Mid Values ($x$)** | $f$ | $dx$ | $fdx$ |
|---|---|---|---|---|
| 0-10 | 5 | 5 | -30 | -150 |
| 10-20 | 15 | 7 | -20 | -140 |
| 20-30 | 25 | 17 | -10 | -170 |
| 30-40 | 35 | 12 | 0 | 0 |
| 40-50 | 45 | 5 | 10 | 50 |
| 50-60 | 55 | 2 | 20 | 40 |
| 60-70 | 65 | 7 | 30 | 210 |
| **Total** | | **55** | | **-160** |

$\bar{x} = A + \frac{\Sigma(fdx)}{N}$ $\bar{x} = 35 + \frac{-160}{55} = 35 - 2.909 = 32.09$

### iii.Step Deviation Method

This technique is an extension of 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:

$$\bar{x} = A + \frac{\Sigma(f \times dx')}{N} \times C$$

Where,

$$dx' = \frac{dx}{C}$$

$C$ – Common factor by which each of the deviation is divided.

$dx = x - A$

**Illustration 3.2.8**

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** |

$\bar{x} = 25 + \frac{34}{80} \times 10$

$= 25 + \frac{340}{80}$

= 25 + 4.25

= 29.25

**Illustration 3.2.9**

The frequency distribution given below gives the cost of production of sugar cane in different holdings. Find the mean.

| Cost in ₹ | 0-20 | 20-40 | 40-60 | 60-80 | 80-100 |
|---|---|---|---|---|---|
| Frequency | 41 | 51 | 64 | 38 | 7 |

**Solution**

**Assumed mean = 50**

| Cost in ₹ | 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** |

$$\bar{x} = A + \frac{\Sigma(f \times dx')}{N} \times C$$

$= 50 - \frac{81}{201} \times 20$

= 50 – 8.06

= 41.94

### 3.2.2.1 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 3.2.10**

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 3.2.11**

The average weight of 150 students in a class is 60 kilograms. 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, $\bar{x}_1 = 70$kgs
Mean weight of girls, $\bar{x}_2 = 55$kgs
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.

**Illustration 3.2.12**

The mean monthly salary paid to all employees in a certain company was ₹600. The mean monthly salaries paid to male and female employees were ₹620 and ₹520 respectively. Find the percentage of male to female employees in the company.

**Solution**

Combined mean, $\bar{x}_{12} = 600$
Mean monthly salaries paid to male, $\bar{x}_1 = 620$
Mean monthly salaries paid to female, $\bar{x}_2 = 520$
Total number of students = $n_1 + n_2$

$$\bar{x}_{12} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

$$600 = \frac{n_1 \times 620 + n_2 \times 520}{n_1 + n_2}$$

$$600(n_1 + n_2) = n_1 \times 620 + n_2 \times 520$$

$$80 n_2 = 20 n_1$$

$$\frac{n_1}{n_2} = \frac{80}{20} = \frac{4}{1}$$

$$n_1 : n_2 = 4 : 1$$

Percentage of male employees = $\frac{4}{5} \times 100 = 80\%$

Percentage of female employees = $\frac{1}{5} \times 100 = 20\%$

### 3.2.2.2 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 3.2.13**

The mean wage of 120 factory workers was found to be ₹17000. It was then discovered that an amount of ₹18750 wage was misread as ₹17850. Find the right mean.

**Solution**

$\overline{x} = \frac{\sum x}{n}$

$17000 = \frac{\sum x}{120}$

$\sum x = 17000 \times 120 = 2040000$

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 3.2.14**

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**

$\overline{x} = \frac{\sum x}{n}$

$45 = \frac{\sum x}{140}$

$\sum x = 140 \times 45 = 6300$

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 $= \frac{\text{Correct} \sum x}{n}$

$= \frac{6381}{140}$

$= 45.58$

### 3.2.3 Geometric Mean

The geometric mean is the n^th 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:

$$GM = \text{Antilog of } \frac{1}{N} (\log x_1 + \log x_2 + \cdots + \log x_n)$$

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 understand.

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.

**A. For individual series**

$$GM = \text{Antilog of } \frac{\sum \log x}{N}$$

**Illustration 3.2.15**

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

$$GM = \text{Antilog of } \frac{\sum \log x}{N}$$

$= \text{Antilog of } \frac{17.6373}{10}$

$= \text{Antilog of } 1.76373$

$= 58.04$

**Illustration 3.2.16**

Calculate the geometric mean of the following data

475, 75, 5, 0.8 0.08, 0.005

**Solution**

| x | Log x |
|---|---|
| 475 | 2.6767 |
| 75 | 1.8751 |
| 5 | 0.6990 |
| 0.8 | $\bar{1}.9031$ |
| 0.008 | $\bar{2}.9031$ |
| 0.005 | $\bar{3}.6990$ |

$\sum \log x$ = 2.6767 + 1.8751+ 0.6990 -1 + 0.9031 – 2 + 0.9031 -3 +0.6990

= 1.756

$GM = \text{Antilog of } \frac{\sum \log x}{N}$

$= \text{Antilog of } \frac{1.756}{6}$

= 1.1962

**B. For discrete series**

$GM = \text{Antilog of } \frac{\sum f \times \log x}{N}$

Where,

x – Value of the variable

N – Number of items

f – Frequency

**Illustration 3.2.17**

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

**C. For continuous series**

GM = Antilog of $\frac{\Sigma f \log x}{N}$

Where,

x – Mid value

**Illustration 3.2.18**

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 $= \text{Antilog of } \frac{\sum f \log x}{N}$

$= \text{Antilog of } \frac{23.2739}{20}$

= Antilog of 1.1637

= 14.58

Income is shown in thousands. Therefore, the Geometric Mean = 14580

**Illustration 3.2.19**

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 | Mid x | log x | f log x |
|---|---|---|---|---|
| 0-10 | 5 | 5 | 0.6990 | 3.4950 |
| 10-20 | 7 | 15 | 1.1761 | 8.2327 |
| 20-30 | 15 | 25 | 1.3979 | 20.9685 |
| 30-40 | 25 | 35 | 1.5441 | 38.6025 |
| 40-50 | 8 | 45 | 1.6532 | 13.2256 |
| **Total** | **60** | | | **84.5243** |

GM $= \text{Antilog of } \frac{\sum f \log x}{N}$

$= \text{Antilog of } \frac{84.5243}{60}$

= Antilog of 1.4087

= 25.62

## 3.2.4 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".

The arithmetic mean of the reciprocal of N values $x_1, x_2, x_3,$ -------$x_n$ is given by

$$\frac{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} \dots\dots\dots \frac{1}{x_n}}{N}$$

The Harmonic Mean is the reciprocal of the arithmetic mean

$$HM = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} \dots\dots\dots \frac{1}{x_n}}$$

$$\therefore HM = \text{Reciprocal of } \left(\frac{\sum \frac{1}{x}}{N}\right) = \frac{N}{\sum \frac{1}{x}}$$

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 understand.
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.

**A. For individual observation**

$$H.M = \frac{N}{\sum\frac{1}{x}}$$

**Illustration 3.2.20**

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 |
| | $\sum\frac{1}{x}$ = **0.27637** |

$$H.M = \frac{N}{\sum \frac{1}{x}}$$

$$H.M = \frac{5}{0.27637} = 18.09$$

Average speed = 18.09 Km/hr

**B. For discrete observation**

$$H.M = \frac{N}{\sum f \frac{1}{x}}$$

**Illustration 3.2.21**

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** |

$$H.M = \frac{N}{\sum f \frac{1}{x}}$$

$$= \frac{26}{2.1081}$$

=12.33

**C. For continuous observations**

$$H.M = \frac{N}{\sum f \frac{1}{x}}$$

Where,

x – Mid Value of the classes

**Illustration 3.2.22**

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

## Recap

- Arithmetic mean - sum of all observations' values divided by the number of observations.
- Sum of deviations of items from the arithmetic mean is always equal to zero.
- Arithmetic mean - best average for all the situations where there are no extreme values in the data.
- Geometric Mean- appropriate average to use for estimating average percent increases in sales, population, production, and so on.
- Geometric Mean- one of the most effective averages for constructing index numbers.
- Harmonic Mean - average used for calculating a company's average rate of profit growth, determining the average speed of a journey, or determining the average price at which items are sold.

## Objective Questions

1. Which measure of central tendency is used for calculating the average speed of a vehicle?
2. What is the average with the limited algebraic treatment?
3. What is the most widely used measure for central tendency?
4. Identify the average in which a figure is extracted from a particular series using mathematical procedures to represent the entire series.
5. When the figures of deviations appear to be large, what method do you use to calculate the Arithmetic Mean?
6. What happens to the arithmetic mean when all the numbers of observations in the data increase by five?
7. If the mean of 12, 11, 9, x, 7 and 5 is 9, find the value of x.
8. What does arithmetic mean when all the data observations decrease by two?
9. Which average is "The $n^{th}$ root of the product of "n" positive values"?

## Answers

1. Harmonic Mean
2. Harmonic mean
3. Arithmetic mean
4. Mathematical average
5. Step deviation method
6. Increase by 5
7. Ten
8. Decrease by 2.
9. Geometric Mean

## Self-Assessment Questions

1. What is arithmetic average? State the advantages and limitations.
2. Explain the concept of combined arithmetic mean.
3. Write a short note on the concept of combined arithmetic mean.
4. What is geometric mean?
5. List 3 situations where harmonic mean can be used.
6. What is harmonic mean?

## Assignments

1. The rain fall in the district for the last five years is, 135, 120, 142, 110 and 150. Calculate the average rainfall of the district for the first five years.

**Ans. 131.4 cm**

2. The following information shows the number of traffic accidents that occurred in 520 cities. Calculate the mean

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| No of accident: | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 |
| No of city: | 38 | 104 | 140 | 78 | 48 | 42 | 28 | 24 | 16 | 2 |

**Ans. 7.78**

3. The following is a list of 200 people's weekly earnings. Determine the mean.

| Weekly wages: | 1000-1200 | 1200-1400 | 1400-1600 | 1600-1800 | 1800-2000 | 2000-2200 | 2200-2400 | 2400-2600 |
|---|---|---|---|---|---|---|---|---|
| No of Persons: | 3 | 21 | 35 | 57 | 40 | 24 | 14 | 6 |

**Ans. 1768**

4. Find Geometric mean

| x: | 18 | 16 | 22 | 12 |
|---|---|---|---|---|

**Ans. 16.61**

5. Speed of four cycles are as follows

| x: | 3.5 | 5.2 | 4.8 | 6.1 | 4.1 |
|---|---|---|---|---|---|

Calculate Harmonic Mean.

**Ans. 4.484**

6. Calculate Geometric mean

| Class: | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 |
|---|---|---|---|---|---|
| Frequency: | 8 | 15 | 25 | 6 | 16 |

**Ans. 21.903**

7. Marks obtained from a class test is given below.

| Mark: | 5-10 | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 |
|---|---|---|---|---|---|---|
| No of students: | 2 | 9 | 29 | 54 | 11 | 5 |

Find Harmonic mean

**Ans. 19.78**

## Suggested Readings


1. Dixon, W.S. and F.J. Massey. (1951). *Introduction to Statistical Analysis*, McGraw Hill Book Company, Inc. New York.
2. Fisher, R.A. and F. Yates. (1948). *Statistical Tables for Biological, Agricultural and Medical Research*, 3rd Ed. Haftner Publishing Company, New York.
3. Sancheti and V.K. Kapoor. (2014). *Business Mathematics. Sultan Chand & Sons*
4. Wikes, F.M (1998*). Mathematics for Business, Finance and Economics*. Thomson Learning

# Unit - 3
# Positional Averages

## Learning Outcomes

After completing this unit, learners will be able to:

- ✓ familiarise with the various types of positional average
- ✓ develop an understanding of which type of average would be the most useful in a particular situation
- ✓ solve problems of median, quartile, decile, percentile, and mode

## Prerequisites

According to Census 2011 statistics produced by Bengaluru - based think tank Takshashila Institution, the populations of Uttar Pradesh and Bihar have the lowest median age - or youngest populations - in India, while Kerala and Tamil Nadu have the highest median age. The median age is the age at which the population is divided in half, with the same number of people older than the median age as under it. A low median age would indicate that a nation has a younger-than-older population. This is the concept where median applies. Therefore, in some situations, positional averages are chosen based on its position within a set of values to effectively represent the data.

The mean may not be appropriate for averaging in certain situations, especially when extreme values significantly impact the data. In such cases, it becomes necessary to utilise positional averages such as the Median, Mode, Quartile, Decile, or Percentile to provide a more representative measure.

## Keywords

Median, Mode, Quartile, Decile, Percentile

## Discussion

### 3.3.1 Positional Average

The term "positional average" refers to an average calculated from a set of observations in which one number based on its position is chosen to represent the complete set. The positional averages discussed in this unit are listed below.

a. Median
b. Quartiles
c. Deciles
d. Percentiles
e. Mode

### 3.3.2 Median

When data is arranged in ascending or descending order, the median is the value of the data's middle most observations.

Median may be defined as "that value of the variable which divides the group into two equal parts, one part comprising all the values greater and other, all values being less than the median". The number of data points affects how the median is calculated. The middle value is the median for an odd number of data, and the average of the two middle values is the median for an even amount of data.

**Advantages of Median**

i. It is rigidly defined.
ii. It is simple to calculate.
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.

**Disadvantages of Median**

i. Median is not suitable for further mathematical treatments
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, and does not take into account every single item in the distribution.

**Computation of Median**

**1. For individual series**

**Steps for calculation**

i. Sort the data into ascending or descending order.
ii. Use the formula.

Median = $(\frac{n+1}{2})^{th}$ item

**Illustration 3.3.1**

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 = $(\frac{n+1}{2})^{th}$ item

= $(\frac{5+1}{2})^{th}$ item

= $3^{rd}$ item

The $3^{rd}$ item in the series is 35.

∴ Median mark is 35

**Illustration 3.3.2**

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

= $3.5^{th}$ 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 $3^{rd}$ and $4^{th}$ item

$$= \frac{12000+15000}{2}$$

$$= \frac{27000}{2}$$

= 13500

Median income of the family is 13500.

## 2. For discrete series

**Steps for calculation**

i. Arrange the data in ascending or descending order
ii. Calculate cumulative frequency (cf)
iii. Determine $\frac{N+1}{2}$
  Where N is the total frequency
iv. Median is the value for the $\left(\frac{N+1}{2}\right)^{th}$ item of the data

**Illustration 3.3.3**

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 = $\frac{N+1}{2}$

= $\frac{115+1}{2}$

= $\frac{116}{2}$

= $58^{th}$ item

∴ Median is the value in the data which comes in the $58^{th}$ 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 $75^{th}$ position,

∴ Median = 900

### 3. For continuous series

**Steps for calculation**

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 3.3.4**

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

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

$$= 30 + \frac{40 - 35}{19} \times 10$$

$$= 30 + \frac{50}{19}$$

$$= 30 + 2.63$$

$$= 32.631$$

**Illustration 3.3.5**

Find the median wage of the following distribution

| Wages (₹): | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|
| No. of labourers: | 3 | 5 | 20 | 10 | 5 |

**Solution**

| Class | f | cf |
|---|---|---|
| 20-30 | 3 | 3 |
| 30-40 | 5 | 8 |
| 40-50 | 20 | 28 |
| 50-60 | 10 | 38 |
| 60-70 | 5 | 43 |
| | N = 43 | |

$$\frac{N}{2} = \frac{43}{2} = 21.5$$

The class having cumulative frequency 21.5 is 40-50

∴ Median class is 40-50

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

$$= 40 + \frac{(21.5 - 8)}{20} \times 10$$

$= 40 + \frac{135}{20}$

$= 40 + 6.75$

$= 46.75$

**Illustration 3.3.6**

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 3.3.7**

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.
For this 0.5 is subtracted from the lower limit and 0.5 is added to the upper 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

Median $= l + \frac{\frac{N}{2} - m}{f} \times c$

$= 19.5 + \frac{(40 - 13)}{28} \times 10$

$= 19.5 + \frac{270}{28}$

= 19.5 + 9.64

= 29.14

### 3.3.3 Quartiles

The values that divide the whole data into four (4) equal parts are called quartiles, when the observations are organised in ascending order. These values are denoted by Q1, Q2 and Q3. It is to be noted that 25% of the data falls below Q1, 50% of the data falls below Q2 and 75% of the data falls below Q3.

Quartile Deviation $Q = \frac{Q_3 - Q_1}{2}$

**Calculation of Quartiles**

**1. For individual series**

**Steps for calculation**

i. Arrange the data in ascending or descending order
ii. Apply the formulas for calculating $Q_1$ and $Q_3$

$Q_1 = (\frac{n+1}{4})^{th}$ item in the series

$Q_3 = (\frac{3(n+1)}{4})^{th}$ item in the series

**Illustration 3.3.8**

The following are the prices of a drug in seven different medical stores.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Price: | 350 | 300 | 425 | 450 | 400 | 600 | 500 |

Find first and third quartiles.

**Solution**

Arrange the data in ascending order

| | | | | | | |
|---|---|---|---|---|---|---|
| 300 | 350 | 400 | 425 | 450 | 500 | 600 |

n = 7

$Q_1 = (\frac{n+1}{4})^{th}$ item in the series

$= (\frac{7+1}{4})^{th}$ item in the series

$= 2^{nd}$ item

= 350

$Q_3 = (\frac{3(n+1)}{4})^{th}$ item in the series

$= (\frac{3(7+1)}{4})^{th}$ item in the series

$= 6^{th}$ item

= 500

**Illustration 3.3.9**

The following table shows the scores of 13 students in a class.

| Marks: | 9 | 13 | 14 | 12 | 17 | 7 | 10 | 8 | 20 | 15 | 21 | 6 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

Calculate lower and upper quartiles

**Solutions**

Arrange the data in ascending order

| Marks: | 6 | 7 | 8 | 9 | 10 | 12 | 13 | 14 | 15 | 17 | 18 | 20 | 21 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

n = 13

$Q_1 = (\frac{13+1}{4})^{th}$ item in the series

$= (\frac{14}{4})^{th}$ item in the series

$= 3.5^{th}$ item

= Value of $3^{rd}$ item + 0.5 (value of $4^{th}$ item – value of $3^{rd}$ item)

= 8 + 0.5 (9-8)

= 8.5

$Q_3 = (\frac{3(13+1)}{4})^{th}$ item in the series

$= 3 \times 3.5^{th}$ item in the series

= 10.5th item

= value of 10th item + 0.5 (value of 11th item – value of 10th item)

=17 + 0.5 (18-17)

= 17.5

**2. For discrete series**

**Steps for calculation**

i. Prepare cumulative frequency table.

ii. Apply the following formula

$Q_1$ = Observation having cumulative frequency $\left(\frac{N+1}{4}\right)$

$Q_3$ = Observation having cumulative frequency $\left(\frac{3(N+1)}{4}\right)$

**Illustration 3.3.10**

Below are the weights of 49 people.

| Weight (in kg): | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 |
|---|---|---|---|---|---|---|---|---|---|
| No. of persons: | 2 | 3 | 6 | 15 | 10 | 5 | 4 | 3 | 1 |

Calculate first and third quartiles.

**Solutions**

| Weight | Frequency | cf |
|---|---|---|
| 58 | 2 | 2 |
| 59 | 3 | 5 |
| 60 | 6 | 11 |
| 61 | 15 | 26 |
| 62 | 10 | 36 |
| 63 | 5 | 41 |
| 64 | 4 | 45 |
| 65 | 3 | 48 |
| 66 | 1 | 49 |
| | **49** | |

N = 49

$Q_1$ = Series having cf $\left(\frac{49+1}{4}\right)$

= Series having cf 12.5

= 61

$Q_3$ = $\left(\frac{3(49+1)}{4}\right)$

= Series having cf 37.5

= 63

**3. For continuous series**

**Steps for calculation**

i. Prepare cumulative frequency table

ii. Identify the class having cumulative frequencies

$Q_1$ class = $\frac{N}{4}$

$Q_3$ class = $\frac{3N}{4}$

iii. Apply the given formula for calculating $Q_1$ and $Q_3$

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - m_1\right)}{f_1} \times c_1$$

$$Q_3 = l_3 + \frac{\left(\frac{3N}{4} - m_3\right)}{f_3} \times c_3$$

Where,

$l_1$ and $l_3$ are the lower classes of quartile classes.

$f_1$ and $f_3$ are the frequencies of the quartile class.

$c_1$ and $c_3$ are the class interval of the quartile class

$m_1$ and $m_3$ are the cumulative frequencies preceding the quartile class

**Illustration 3.3.11**

Sales of 100 companies are given below. Calculate first and third quartiles.

| Sales (in'000): | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 |
|---|---|---|---|---|---|---|---|---|
| No. of days: | 13 | 17 | 50 | 60 | 55 | 45 | 23 | 7 |

**Solution**

| x | f | cf |
|---|---|---|
| 0-10 | 13 | 13 |
| 10-20 | 17 | 30 |
| 20-30 | 50 | 80 |
| 30-40 | 60 | 140 |
| 40-50 | 55 | 195 |
| 50-60 | 45 | 240 |
| 60-70 | 23 | 263 |
| 70-80 | 7 | 270 |
| | **N = 270** | |

$Q_1$ Class = $\left(\frac{N}{4}\right)^{th}$ Class

= $\left(\frac{270}{4}\right)^{th}$ Class

= $67.5^{th}$ Class

= 20-30

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - m_1\right)}{f_1} \times c_1$$

$= 20 + \frac{(67.5-30)}{50} \times 10$

$= 20 + \frac{375}{50}$

= 20 + 7.5

= 27.5

Q3 Class = $\left(\frac{3N}{4}\right)^{th}$ Class

= $3 \times 67.5^{th}$ class

= $202.5^{th}$ class

= 50-60

$$Q_3 = l_3 + \frac{\left(\frac{3N}{4} - m_3\right)}{f_3} \times c_3$$

$$= 50 + \frac{(202.5 - 195)}{45} \times 10$$

$$= 50 + \frac{75}{45}$$

$$= 50 + 1.67$$

$$= 51.67$$

The above result shows that $Q_1 = 27.5$ and $Q_3 = 51.67$

**Illustration 3.3.12**

Calculate the quartile deviation for the following data of annual income of 100 families.

| Annual Income: | Less than 499 | 500-999 | 1000-1999 | 2000-2999 | Above 3000 |
|---|---|---|---|---|---|
| No. of families: | 5 | 25 | 40 | 20 | 10 |

| Ann. Income | Actual class | f | cf |
|---|---|---|---|
| Less than 499 | Less than 499.5 | 5 | 5 |
| 500-999 | 499.5-999.5 | 25 | 30 |
| 1000-1999 | 999.5-1999.5 | 40 | 70 |
| 2000-2999 | 1999.5-2999.5 | 20 | 90 |
| Above 3000 | Above 2999.5- | 10 | 100 |
| | | **N = 100** | |

$Q_1$ Class = $\left(\frac{N}{4}\right)^{th}$ Class

= $\left(\frac{100}{4}\right)^{th}$ Class

= $25^{th}$ Class

= 500-999 class

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - m_1\right)}{f_1} \times c_1$$

$$= 499.5 + \frac{(25-5)}{25} \times 500$$

$$= 499.5 + 4$$

$= 899.5$

Q3 Class = $\left(\frac{3N}{4}\right)^{th}$ Class

$= 75^{th}$ class

= 2000-2999 class

$$Q_3 = l_3 + \frac{\left(\frac{3N}{4} - m_3\right)}{f_3} \times c_3$$

$$= 1999.5 + \frac{(75-70)}{20} \times 1000$$

$= 1999.5 + 250$

$= 2249.5$

$$Q = \frac{Q_3 - Q_1}{2}$$

$$= \frac{2249.5 - 899.5}{2}$$

$$= \frac{1350}{2} = 675$$

## 3.3.4 Deciles

The values that divide the complete data into ten (10) equal parts are called deciles when the observations are organised in ascending order. These values are denoted by D1, D2,…, D9. It is to be noted that 10% of the data falls below D1, 20% of the data falls below D2,…,, and 90% of the data falls below D9.

## 3.3.5 Percentiles

The values that divide the whole data into hundred (100) equal sections are called percentiles when the observations are organised in ascending order. These values are denoted by P1, P2,…, P99. It is to be noted that 1% of the data falls below P1, 2% of the data falls below P2,…., and 99% of the data falls below P99.

## 3.3.6 Mode

Mode of a data 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.

**Illustration 3.3.13**

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 3.3.14**

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 3.3.15**

The following table shows the monthly income of 148 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 |

**Solution**

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) | Grouping in twos (3) | Grouping in threes (4) | Grouping in threes (5) | Grouping in threes (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, which 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 for calculation**

a. Locate the modal class having highest frequency or by preparing analysis table.
b. Apply this formula

$$\text{Mode} = l + \frac{(f_1 - f_0)}{2f_1 - f_0 - f_2} \times c$$

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 3.3.16**

The following table shows the frequency distribution of the marks of 72 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

$$\text{Mode} = l + \frac{(f_1 - f_0)}{2f_1 - f_0 - f_2} \times c$$

$$= 50 + \frac{(15-12)}{2*15-12-14} \times 10$$

$$= 50 + \frac{30}{4}$$

= 57.5

**Illustration 3.3.17**

The following table shows the monthly income of 125 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 | Grouping in twos | | Grouping in threes | | |
|---|---|---|---|---|---|---|
| | (1) | (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 / F column | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 |
|---|---|---|---|---|---|---|---|---|
| 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.

$$\text{Mode} = l + \frac{(f_1 - f_0)}{2f_1 - f_0 - f_2} \times c$$

l =30

$f_1 = 30$

$f_0 = 18$

$f_2 = 20$

c = 10

$$= 30 + \frac{(30-18)}{(2 \times 30) - 18 - 20} \times 10$$

$$= 30 + \frac{120}{22}$$

= 30 + 5.45

= 35.45

**Calculation of mode by Graphical Method**

In graphical method, the series is represented by the rectangular diagram. The value of the mode can be determined graphically only in continuous series.

The following are the steps of calculating mode by graphical method.

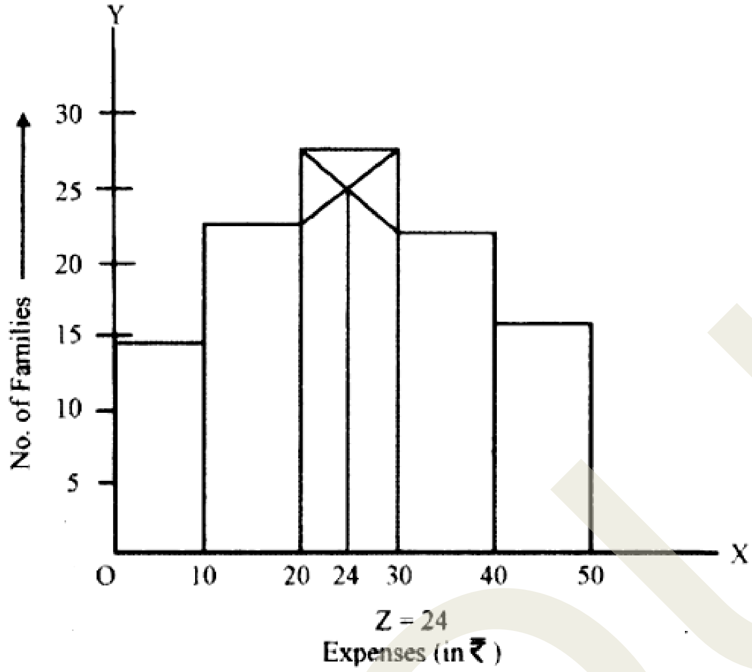
1. According to the given frequency distribution, draw a histogram.
2. The highest rectangle in this diagram is modal-class.
3. Join the top right and left edges of this rectangle with the top right edge of the rectangle representing the preceding class and the top left edge of the rectangle representing the succeeding class respectively.
4. From the point of intersection of both the lines draw a straight line downwards perpendicular on the X-axis. The point where the straight line meets the X-axis gives us the modal – value.

**Illustration 3.3.18**

Find out the mode of the following frequency distribution by the graphical method

| | | | | | |
|---|---|---|---|---|---|
| Expense in ₹: | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 |
| No. of families: | 14 | 23 | 27 | 21 | 15 |

**Solution**




The figure shows the expenditure on X axis and the number of families on the Y axis. The rectangle of 20-30 class is highest. The lines joining both top vertices of the rectangle, with the opposite right and left top vertices of the preceding and succeeding rectangles, intersect one another at the point E. On putting the perpendicular from point E to the X axis, the X axis is intersected on the mode value of the distribution. Therefore Mode = 24.

### 3.3.7 Empirical Relation between median and mode

For moderately asymmetrical distribution (or for asymmetrical curve), the relation
**Mean – Mode = 3 (Mean - Median),**
approximately holds. In such a case, first evaluate mean and median. Mode is determined by
**Mode = 3 Median – 2 Mean.**

## Recap

- Median - middle number in a list of numbers sorted in ascending or descending order.
- Mode- most often occurring value in a series.
- Quartiles- values that divide the whole data into four equal parts.
- Deciles- values that divide the complete data into ten equal parts.
- Percentiles- values that divide the whole data into hundred equal sections.
- Median is sometimes used as opposed to the mean when there are outliers in the sequence that might affect the average of the values.

## Objective Questions

1. What is the name of an average calculated from a set of observations in which one number based on its position is chosen to represent the complete set?
2. Which positional average divides the whole data into hundred equal sections?
3. Explain the link between the mean, median, and mode?
4. Into how many parts does deciles divide the whole data?
5. Name the average that is commonly referred to as the "middle most value of observation."
6. Which positional average divides the whole data into four equal parts?
7. Name the average, often known as the "most recurring value in a series."

## Answers

1. Positional average
2. Percentile
3. Mode = 3 Median -2 Mean
4. 10
5. Median
6. Quartile
7. Mode

## Self-Assessment Questions

1. Explain the relationship between mean, median and mode.
2. What is a median?
3. Differentiate between quartiles and deciles.
4. State the merits and demerits of using mode.
5. What do you mean by a positional average? Explain any 2 positional averages.

## Assignments

1. Find out the median of the following observation

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| X: | 46 | 64 | 87 | 41 | 58 | 35 | 77 | 55 | 90 | 33 | 92 |

If 92 is replaced with 99 and 41 with 43 in the above data, find the new median.

**Ans.58**

2. The following table represents the income of 122 families. Calculate Median income.

| | | | | | | |
|---|---|---|---|---|---|---|
| Income: | 1000 | 1500 | 3000 | 2000 | 2500 | 1800 |
| No. of family: | 24 | 26 | 16 | 20 | 6 | 30 |

**Ans. 1800**

3. Run scored by a batter in the cricket tournament are as follows. Calculate the modal run.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Runs: | 16 | 16 | 14 | 22 | 14 | 13 | 15 | 24 | 12 | 23 | 14 | 20 | 17 | 21 | 14 |

**Ans. 14**

4. 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 |

**Ans. 40**

5. Find mode from the following data.

| | | | | | | |
|---|---|---|---|---|---|---|
| Mark: | 20-24 | 25-29 | 30-34 | 35-39 | 40- 44 | 45-49 |
| No. of student: | 20 | 24 | 32 | 28 | 20 | 26 |

**Ans. 32.83**

6. Ages of members in a joint family are as follows

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Age: | 38 | 7 | 43 | 15 | 20 | 25 | 12 | 18 | 11 |

Calculate quartiles.

**Ans.** $Q_1 = 11.5, Q_2 = 18, Q_3 = 31.5$

7. The following table depicts the number of road accident occurred in a city. Find quartiles.

| | | | | | |
|---|---|---|---|---|---|
| No of Road Accident: | 20 | 30 | 40 | 50 | 60 |
| No of cities: | 4 | 16 | 20 | 18 | 11 |

**Ans.** $Q_1 = 30, Q_2 = 40, Q_3 = 50$

## Suggested Readings


1. Gupta, S.C. and Kapoor, V.K. *Fundamentals of Mathematical Statistics*, Sultan Chand and Sons, Delhi, XI edition, 2002.
2. Gupta, S.C. and Kapoor, V.K. *Fundamentals of Applied Statistics*. Sultan Chand and Sons, Delhi, 4th Edition, January 2007.
3. Dr. P.R. Vittal (2012). *Business Maths & Statistics*. Margham Publications
4. C.R.Kothari (2013). *Quantitative Techniques*. Vikas Publishing House
5. S.P.Guptha-(2012). *Statistical Methods*. Sultan Chand &Sons,New Delhi

4

BLOCK

# Measures of Dispersion

# Unit - 1
# Dispersion

## Learning Outcomes

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

- ✓ familiarise the term dispersion
- ✓ identify the uses of dispersion
- ✓ differentiate between absolute measures of dispersion and relative measures of dispersions

## Prerequisites

Suppose a teacher tabulates the marks scored by students of two classes. Both classes have the same average score in statistics, which is 75. However, the scores within each class are very different:

- **Class A:** Scores are 70, 72, 74, 76, 78, and 80.
- **Class B:** Scores are 50, 60, 70, 80, 90, and 100.

While the average score is the same for both classes, the distribution of scores in Class A is much more consistent, with all scores close to the average (low dispersion). In contrast, Class B has a much wider range of scores, with some students performing very well and others performing poorly (high dispersion). Dispersion helps understand the variability in data, even when averages are identical. Here, it would indicate that Class A has more uniform performance, while Class B has significant variation.

## Keywords

Absolute Measures of Dispersion - Relative Measures of Dispersion

# Discussion

### 4.1.1 Introduction

Imagine you are a teacher and you want to compare the performance of two classes in a recent test. You have the test scores of each student in both classes. To get an overall idea of how well the students performed, you first calculate the measures of central tendency for each class. You find that Class A has a mean score of 85, and Class B has a mean score of 78. These mean scores give you a general sense of the typical performance in each class.

Now, you want to understand how the individual test scores are spread out around these mean scores. This is where measures of dispersion come into play. Measures of dispersion help you see if the scores in each class are tightly clustered around the mean or if they are widely scattered.

Let's consider two scenarios for each class:

***Scenario 1 (Class A):***

In Class A, the test scores are relatively close to the mean of 85. Most students scored between 82 and 88, with only a few outliers scoring much higher or lower. In this case, the measures of dispersion indicate that the scores are tightly clustered around the mean, suggesting a consistent performance among the students.

***Scenario 2 (Class B):***

In Class B, the test scores show more variation. Some students scored as low as 65, while a few others scored as high as 95. The majority of scores are spread between 70 and 85. In this case, the measures of dispersion indicate a wider spread of scores around the mean of 78, suggesting a more diverse performance among the students.

By comparing the measures of dispersion between the two classes, you can get a better understanding of how the student's performances differ. Class A shows a more consistent performance, with scores clustered closely around the mean, while Class B demonstrates a wider range of scores, indicating a greater variation in performance.

Based on the above example it is clear that the measures of dispersion complement measures of central tendency by providing insights into the variability and spread of data points. They allow you to gauge how individual values deviate from the central values, giving a more complete picture of the dataset's characteristics.

## 4.1.2 Meaning and Definition

The averages give an understanding of the distribution's central tendency, but it's also important to understand how the variables are clustered around or spread away from the average. The term "dispersion" refers to the degree to which variables deviate from its average. The degree of scatter or variations of the variable around a central value is referred to as dispersion.

According to A.L Bowley, "Dispersion is the measure of variation of the items."

According to Spiegel, "The degree to which numerical data tend to spread about an average value is called the variation or dispersion of the data."

## 4.1.3 Purpose of Dispersion

The objectives of measures of dispersion are diverse and serve various purposes in statistical analysis. It includes:

i. ***Comparing Variability:*** Measures of dispersion allow us to compare the variability between two or more series of data. By examining the spread or dispersion of values, we can determine which series has more or less variation.

ii. ***Controlling Variability:*** Understanding the measures of dispersion helps in controlling variability. By analysing and identifying sources of variation, we can implement strategies or interventions to reduce or manage the variability in a given dataset.

iii. ***Assessing Consistency of Measures of Central Tendency:*** Measures of dispersion provide insights into the consistency of measures of central tendency, such as the mean or median. By examining the dispersion around these central values, we can assess the stability and reliability of these measures.

iv. ***Comparing and Contrasting Series:*** Measures of dispersion facilitate the comparison and contrast of two or more series in terms of their consistency. By evaluating the dispersion, we can determine which series has more consistent or similar values and identify any notable differences.

v. ***Examining Deviation from Central Value:*** Measures of dispersion demonstrate how far individual items in a series deviate from the central value, such as the mean. This information helps in understanding the extent of individual variation from the average or central tendency.

vi. ***Facilitating Additional Statistical Measures:*** Measures of dispersion play a crucial role in conducting various advanced statistical analyses. They provide necessary information for applying statistical techniques like correlation, regression, hypothesis testing, and other analyses that rely on understanding the spread and variability of data points.

vii. ***Supporting Advanced Statistical Analysis:*** Measures of dispersion are vital in advanced statistical analysis. They enhance our understanding of the data distribution, aid in detecting outliers or extreme values, and contribute to making accurate statistical inferences.

### 4.1.4 Properties of dispersion

The properties of measures of dispersion are as follows:

i. ***Statistical Information:*** Measures of dispersion are concerned with statistical information and provide valuable insights into the spread, variability, and distribution of data points within a dataset.

ii. ***Well-Defined:*** Measures of dispersion must be well defined and have clear and unambiguous definitions and calculations. They should follow established mathematical principles and be consistently applicable.

iii. ***Reliability of an Average:*** Measures of dispersion demonstrate the reliability of an average or a measure of central tendency. By examining the dispersion around the average, we can assess the stability and trustworthiness of the central value.

iv. ***Based on All Observations:*** A good measure of dispersion should be based on all observations in the dataset. It should consider the entire range of values and not selectively exclude or ignore certain data points.

v. ***Quantitative Methods:*** Measures of dispersion consist of various quantitative methods through which variances can be measured. These methods may include range, variance, standard deviation, interquartile range, or other statistical calculations that capture the spread of data points.

vi. ***Comprehensibility and Calculation:*** Measures of dispersion should be simple to comprehend and calculate. They should not require overly complex or convoluted formulas, making them accessible to a wide range of users.

vii. ***Degree of Divergence:*** Measures of dispersion show the degree to which different items or series diverge from the central or core value, such as the mean. They provide a quantitative understanding of how much individual values deviate from the average or central tendency.

viii. ***Algebraic Treatment:*** Measures of dispersion should be capable of algebraic treatment. This means they should be amenable to mathematical operations, allowing for further statistical analyses, computations, and manipulations.

ix. ***Minimal Influence of Extreme Observations:*** A reliable measure of dispersion should be minimally affected by extreme observations or outliers. Extreme values should have little impact on the overall measure of dispersion, ensuring that it accurately represents the typical variability of the dataset.

x. ***Stability against Sampling Fluctuations:*** Measures of dispersion should be stable and not heavily influenced by fluctuations in sampling. They should provide consistent and reliable estimates of the variability, regardless of the specific sample selected from the population.

## 4.1.5 Methods or Types of Dispersion

The following are some of the most common methods of studying dispersion.

i. Range
ii. Inter - quartile range
iii. Quartile deviation
iv. Mean deviation
v. Standard deviation
vi. Lorenz curve

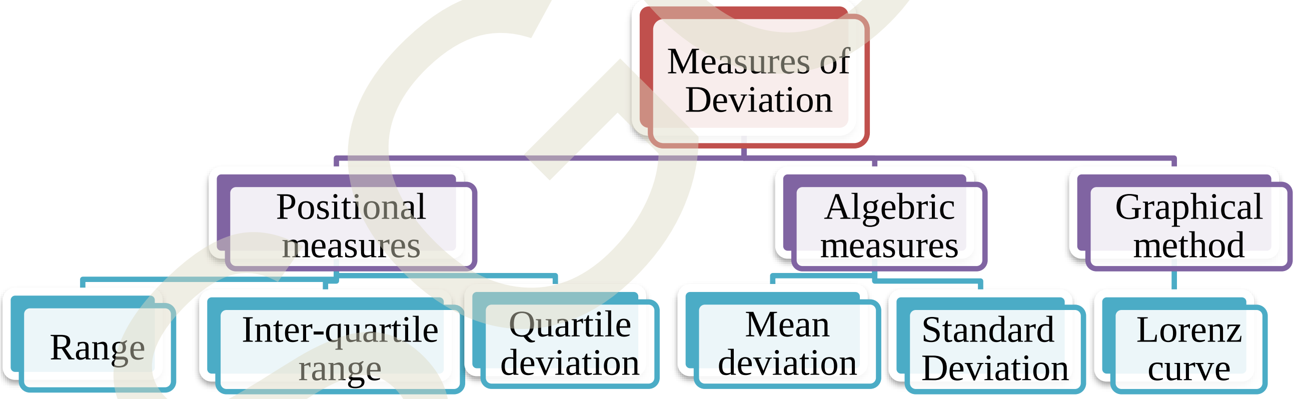



**Fig 4.1.1** Methods or Types of dispersion

## 4.1.6 Absolute and Relative Measures of Dispersion

Absolute and relative measures of dispersion are two approaches used to quantify and compare the spread or variability of data within a dataset. Let's explore each of these measures in more detail:

***a. Absolute Measures of Dispersion:*** Absolute measures of dispersion provide a direct measurement of the spread or variability in the original unit of measurement. These measures give a numerical value that represents the actual range or extent of variation in the data. Some commonly used absolute measures of dispersion include:

i. The Range
ii. The Quartile Deviation
iii. The Mean Deviation
iv. The Standard Deviation

***b. Relative Measures of Dispersion:*** Relative measures of dispersion express the spread or variability of data relative to a reference point, typically a measure of central tendency. These measures provide a way to compare the variability across different datasets, even if they have different scales or units of measurement. Some commonly used relative measures of dispersion include:

i. Relative Range
ii. Relative Quartile Deviation
iii. Relative Mean Deviation
iv. Coefficient of variation

Relative measures of dispersion are particularly useful when comparing datasets with different scales or units, as it provides a standardised way to evaluate the spread relative to the central tendency.

Both absolute and relative measures of dispersion have their advantages and are chosen based on the specific context and requirements of the analysis. Absolute measures provide direct insights into the spread of the data, while relative measures facilitate comparisons and standardisation across different datasets.

## 4.1.7 Difference between absolute and relative measures of dispersion

The difference between absolute and relative measures of dispersions are as follows:

| Aspect | Absolute Measures of Dispersion | Relative Measures of Dispersion |
|---|---|---|
| Definition | Measures that quantify the spread of data in its original units. | Measures that express the spread of data relative to the mean or another reference point. |

| | | |
|---|---|---|
| Calculation | Calculated using the actual values of the data. | Calculated using ratios or percentages. |
| Examples | Range, Mean Deviation, Variance, Standard Deviation | Coefficient of Variation, Relative Range, Coefficient of Mean Deviation |
| Interpretation | Provides information on the spread of data in the original units. | Helps to compare the dispersion of data across different datasets or variables. |
| Unit of Measurement | Retains the same unit as the data. | Unitless or expressed as a percentage. |
| Comparative Analysis | Suitable for comparing datasets with similar units. | Useful for comparing datasets with different units or scales. |
| Mathematical Properties | Can sum up the measures of dispersion. | Measures may not be directly additive or have mathematical properties like additivity. |
| Calculation Complexity | Often simpler to calculate compared to relative measures. | May require additional calculations or transformations. |

## Recap

- Dispersion – measures to quantify data deviation from averages.
- Dispersion - provides insights into the spread or variability of data.
- Common measures of dispersion include range, mean deviation, quartile deviation, and standard deviation.
- Dispersion - indicate data heterogeneity or homogeneity.
- Dispersion measures -crucial in risk assessment, quality control, finance, and decision making.
- Absolute measures of dispersions -quantify original unit spread.
- Absolute measures - influenced by outliers.
- Relative measures of dispersion - allow comparisons regardless of units.

## Objective Questions

1. Which measure of dispersion provides information about the spread of data in its original units?
2. Which measure of dispersion is less affected by outliers?
3. What does dispersion measure in a dataset?
4. Which measure of dispersion provides a direct understanding of variability?
5. Which measure of dispersion allows comparisons regardless of units or scales?

## Answers

1. Absolute measures of dispersion
2. Relative measures of dispersion
3. Spread or variability
4. Absolute measures of dispersion.
5. Relative measures of dispersion

## Self-Assessment Questions

1. Illustrate with examples how measures of dispersion can provide insights into the shape and characteristics of different types of distributions, such as symmetric, skewed, or multimodal distributions.
2. Elaborate on the relationship between measures of central tendency and measures of dispersion in data analysis.
3. Explain relative measures of dispersion and how they facilitate comparisons between datasets.

## Assignments

1. Define measures of dispersion and explain their importance in statistical analysis.
2. Discuss the concept of range as a measure of dispersion. Highlight its strengths and limitations.
3. Compare and contrast the absolute measures of dispersion, such as mean deviation, variance, and standard deviation.
4. Discuss the impact of outliers on absolute measures of dispersion and how relative measures are more robust in such cases.

## Suggested Readings


1. Dr. S M Shukla and Dr. Sahai (2020) - *Principles of statistics*, Sahitya Bhavan Publication, Delhi
2. G S Monga (2000) - *Mathematics and Statistics for economics*, Vikas Publishing House Pvt Ltd.
3. S.C Gupta & V K Kapoor (2020) - *Fundamentals of Mathematical Statistics*, Sultan Chand & Sons Educational publishers
4. S.C. Gupta (2023) - *Fundamentals of Statistics*, Himalaya Publishing House.
5. Frederick E Croxton, Dudley J Cowden, Sidney Klein (2009) - *Applied General Statistics*, Prentice Hall India.
6. Naval Bajpai (2013) - *Business statistics*, Pearson Educational Publications

# Unit - 2

# Measures of Dispersion

## Learning Outcomes

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

- ✓ familiarise themselves with various types of dispersion.
- ✓ calculate the variations in a data set.
- ✓ compare the reliability and consistency of two or more data sets.
- ✓ calculate the combined standard deviation.
- ✓ rectify errors that occur during the computation of standard deviations.

## Prerequisites

Suppose a company wants to evaluate the reliability of two suppliers for delivering materials on time. Over the past 10 days, *Supplier X* has delivered materials with a deviation of ±1 day from the promised delivery time, and *Supplier Y* delivered materials with a deviation ranging from ±5 to ±7 days. If the company looks only at the average delivery time of each supplier, it might assume both are equally reliable. However, by calculating a measure of dispersion (e.g., standard deviation), the company can quantify the variability in delivery times and recognize that Supplier X is more reliable, as their deliveries are more consistent. Measures of dispersion are essential for making informed decisions, as they provide a clearer picture of consistency and variability in data.

## Keywords

Range, Quartile Deviation, Mean Deviation, Standard Deviation, Coefficient of Range, Coefficient of Quartile Deviation, Coefficient of Mean Deviation, Coefficient of Variation

## Discussion

### 4.2.1 Range

The range is defined as the difference between the highest and lowest values in a series, representing the span between two extreme values and indicating the limits within which the values vary.

**Range = L-S**

Where,

L = Largest value

S = Smallest value

Coefficient of Range = $\frac{L-S}{L+S}$

**Advantages of Range**

a) It is the most basic method of determining dispersion.

b) It is simple to comprehend and calculate.

c) It is rigidly defined.

d) Even if some items in the midst of a series are missing, its calculation is unaffected.

**Disadvantages of Range**

a) It prioritises only the two extreme values.

b) On many occasions, it is not a reliable measure of dispersion.

c) It is influenced by sampling fluctuations.

d) It cannot be further algebraically treated.

e) It does not take into account the frequencies of the distribution.

f) Range cannot be found for open ended distributions

**Usage of Range**

Range is useful in the following situations,

a) Range can be utilised as an efficient quality control method.

b) 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.

c) For weather forecasts, the meteorological department uses a range.

d) Range can be applied in areas where the data have small variations

**Computation of Range**

**i. For individual series**

**Range = L - S**

Where,

L = Largest value

S = Smallest value

**Illustration 4.2.1**

Below are the prices of 1 kg of apples for the first six months. Find Range and Coefficient of Range.

| Month: | January | February | March | April | May | June |
|---|---|---|---|---|---|---|
| Price/kg: | 120 | 115 | 150 | 130 | 175 | 160 |

**Solution**

Range = L-S

L =175

S = 115

Range = 175-115

= 60

Coefficient of Range = $\frac{L-S}{L+S}$

$= \frac{175-115}{175+115}$

$= \frac{60}{290}$

= 0.206

**ii. For discrete series**

**Illustration 4.2.2**

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 = $\frac{510-210}{510+210}$

= $\frac{300}{720}$

= 0.416

**iii. For continuous series**

First method

$$Range = Upper\ limit\ of\ the\ highest\ class\ interval(L) - Lower\ limit\ of\ the\ lowest\ class\ interval(S)$$

$$Coefficient\ of\ range = \frac{L-S}{L+S}$$

**Illustration 4.2.3**

From the following data calculate Range and Coefficient of range

| | | | | | | |
|---|---|---|---|---|---|---|
| **Mark:** | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
| **No of Person:** | 14 | 12 | 15 | 8 | 6 | 5 |

**Solution**

Range = L-S

= 70 – 10

= 60

Coefficient of Range = $\frac{70-10}{70+10}$

= $\frac{60}{80}$

= 0.75

**Second method:**

$$Range = Mid\ points\ of\ highest\ interval\ (L) - Mid\ point\ lowest\ interval\ (S)$$

$$Coefficient\ of\ range = \frac{L-S}{L+S}$$

Applying the second method on Illustration 4.2.3:

Range = L – S

= 66 – 16

= 50

Coefficient of range = $\frac{L-S}{L+S}$

= $\frac{66-16}{66+16}$

= $\frac{50}{82}$

= 0 .610

## 4.2.2 Quartile Deviation

The interquartile range is a measure of dispersion based on the upper quartile $Q_3$ and lower quartile Q1. The quartile deviation is a measure of dispersion based on quartiles. It is the half of the difference between the upper and lower quartile. It is obtained by dividing interquartile range by 2. Therefore, it is also known as semi- inter quartile range.

Inter quartile range = $Q_3 - Q_1$

QD = $\frac{Q_3-Q_1}{2}$

Coefficient of Quartile Deviation = $\frac{Q_3-Q_1}{Q_3+Q_1}$

Where,

QD – Quartile deviation

$Q_1$ – First quartile

$Q_3$ – Third quartile

Quartile deviation is a better measure than range since it makes use of 50 per cent of the data compared to range which is based on only the highest and lowest values.

**Advantages of Quartile Deviation**

- It is easy to understand and calculate
- It can be calculated for open-ended classes.
- It is unaffected by extreme values.

**Disadvantages of Quartile Deviation**

- It is not based on all observation
- It is not capable of further algebraic treatment.

**Computation of Quartile Deviation**

**i. For individual series**

**Illustration 4.2.4**

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

$= 3.5^{th}$ item

$= 3^{rd}$ item + 0.5 ($4^{th}$ item $- 3^{rd}$ item)

= 8 + 0.5 (9-8)

= 8.5

$Q_3$ = value of $\frac{3(n+1)}{4}$ th item

= 3 × 3.5th item

= 10.5th item

= 10th item + 0.5 (11th item – 10th item)

= 17 + 0.5 (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

**ii. For discrete series**

**Illustration 4.2.5**

Below are the heights (in inches) of 49 people.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Height (in inches):** | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 |
| **No of persons:** | 2 | 3 | 6 | 15 | 10 | 5 | 4 | 3 | 1 |

Calculate inter-quartile range, quartile deviation, and coefficient of quartile deviation.

**Solution**

| Height (in inches) | Frequency | Cum.f |
|---|---|---|
| 58 | 2 | 2 |
| 59 | 3 | 5 |
| 60 | 6 | 11 |
| 61 | 15 | 26 |
| 62 | 10 | 36 |
| 63 | 5 | 41 |
| 64 | 4 | 45 |
| 65 | 3 | 48 |
| 66 | 1 | 49 |
| | **49** | |

N = 49

$Q_1$ = Series having cf $\left(\frac{49+1}{4}\right)$

= Series having cf 12.5

= 61

$Q_3 = \left(\frac{3(49+1)}{4}\right)$

= Series having cf 37.5

= 63

Inter-quartile range = $Q_3 - Q_1$

= 63-61

= 2

Quartile Deviation = $\frac{Q_3 - Q_1}{2}$

$= \frac{63-61}{2}$

$= \frac{2}{2}$

= 1

Coefficient of Quartile Deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{63-61}{63+61}$

$$= \frac{2}{124}$$

$= 0.016$

iii.**For continuous series**

**Illustration 4.2.6**

The salaries of 270 employees are given below. Calculate the inter-quartile range, quartile deviation, and coefficient of quartile deviation.

| **Salary (in'000):** | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 |
|---|---|---|---|---|---|---|---|---|
| **No of employees:** | 13 | 17 | 50 | 60 | 55 | 45 | 23 | 7 |

**Solution**

| X | f | c.f. |
|---|---|---|
| 0-10 | 13 | 13 |
| 10-20 | 17 | 30 |
| 20-30 | 50 | 80 |
| 30-40 | 60 | 140 |
| 40-50 | 55 | 195 |
| 50-60 | 45 | 240 |
| 60-70 | 23 | 263 |
| 70-80 | 7 | 270 |
| | N = 270 | |

$Q_1$ Class $= \left(\frac{N}{4}\right)^{th}$ Class

$= \left(\frac{270}{4}\right)^{th}$ Class

$= 67.5^{th}$ Class

$= 20\text{-}30$

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - m_1\right)}{f_1} \times c_1$$

$= 20 + \frac{(67.5-30)}{50} \times 10$

$= 20 + \frac{375}{50}$

$= 20 + 7.5$

$= 27.5$

Q3 Class $= 3(\frac{N}{4})^{th}$ Class

$= 3 \times 67.5^{th}$ class

$= 202.5^{th}$ class

= 50-60

$Q_3 = l_3 + \frac{3\left(\frac{N}{4} - m_3\right)}{f_3} \times c_3$

$= 50 + \frac{(202.5-195)}{45} \times 10$

$= 50 + \frac{75}{45}$

$= 50 + 1.67$

$= 51.67$

$Q_1 = 27.5$

$Q_3 = 51.67$

Inter quartile range $= Q_3 - Q_1$

$= 51.67 - 27.5$

$= 24.17$

Quartile Deviation $= \frac{Q_3 - Q_1}{2}$

$= \frac{51.67-27.5}{2}$

$= \frac{24.17}{2}$

$= 12.085$

Coefficient of Quartile Deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{51.67-27.5}{51.67+27.5}$

$$= \frac{24.17}{79.17}$$

$$= 0.305$$

### 4.2.3 Mean Deviation

The arithmetic mean of the absolute deviation of the observations from an assured average is called mean deviation. In other words, the arithmetic average of the deviations of items in a series taken from its central value, ignoring the plus and minus sign, is known as mean deviation. It is also known as Average deviation.

$$\text{Mean deviation} = \frac{\sum|x-A|}{n}$$

Where,

A – Any average. i.e., Mean or Median or Mode

|x-A| is read as modulus (x-A) is the modulus or absolute value of the deviation obtained after ignoring the negative sign.

**Coefficient of Mean Deviation**

Mean deviation is divided by the average to get the coefficient of mean deviation. If deviations are taken from mean, we divide it by mean, if the deviations are taken from median, then it is divided by median and if the deviations are taken from mode, then we divide mean deviation by mode. It is calculated to compare the data of two series.

$$\text{Coefficient of Mean Deviation} = \frac{MD}{A}$$

Where,

MD - Mean Deviation

A - Any average. i.e., Mean or Median or Mode

**Coefficient of Mean variation**

$$\text{Coefficient of Mean variation} = \frac{MD}{A} \times 100$$

**Advantages of mean deviation**

a) It is based on all observation

b) It is rigidly defined

c) Extreme values have little impact on it.

d) It truly represents the average of deviations of the items of a series by removing the irregularities in the distribution.

e) It can be calculated from any value

f) In the realm of business and trade, it is quite useful.

g) It is concise to compute and comprehend

**Disadvantages of mean deviation**

a) It cannot be treated mathematically any further.

b) Ignoring signs of deviations may create artificiality

c) When the actual value of an average is a fraction, it is difficult to calculate.

d) It cannot be used for open end classes.

e) The volatility of sampling has a significant impact on it.

**Computation of mean deviation, coefficient of mean deviation and coefficient of mean variation**

**i. For individual series**

**Steps,**
**For calculating Mean Deviation**

a. Calculate the average (mean, median, or mode) required to calculate the mean deviation.

b. Take the deviation of an item from the average (mean, median, or mode). i.e., $|x - A|$

c. Calculate the total value of the deviation. i.e $\sum|x - A|$

d. Divide the total value of deviation obtained by the number of observations.

$$MD = \frac{\sum|x-A|}{n}$$

**For calculating coefficient of mean deviation**

$$\text{Coefficient of Mean Deviation} = \frac{MD}{A}$$

**For calculating coefficient of mean variation**

$$\text{Coefficient of Mean variation} = \frac{MD}{A} \times 100$$

**Illustration 4.2.7**

From the following data, compute the mean deviation from the mean, the coefficient of mean deviation, and the coefficient of mean variation.

X: 8 11 5 4 5 0 2 6 9 3 2

**Solution**

Mean deviation about mean = $\frac{\sum|x - Mean|}{n}$

Mean = $\frac{\sum x}{n}$

= $\frac{55}{11}$

= 5

| **X** | **$\|x - Mean\|$** |
|---|---|
| 8 | 3 |
| 11 | 6 |
| 5 | 0 |
| 4 | 1 |
| 5 | 0 |
| 0 | 5 |
| 2 | 3 |
| 6 | 1 |
| 9 | 4 |
| 3 | 2 |
| 2 | 3 |
| **$\sum x = 55$** | **$\sum\|x - A\| = 28$** |

Mean deviation = $\frac{\sum|x - Mean|}{n}$

= $\frac{28}{11}$

= 2.545

Coefficient of Mean Deviation = $\frac{\text{MD}}{\text{mean}}$

$= \frac{2.545}{5}$

= 0.509

Coefficient of Mean Variation = $\frac{\text{MD}}{\text{mean}} \times 100$

$= \frac{2.545}{5} \times 100$

= 50.90 %

**Illustration 4.2.8**

The prices of a kilograms of orange in various markets are listed below.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Price: | 120 | 130 | 140 | 110 | 160 | 150 | 190 | 180 | 170 | 200 |

Find

a. Mean deviation about median
b. Coefficient of Mean deviation
c. Coefficient of Mean variation.

**Solution**

Mean deviation about median = $\frac{\sum|x-\text{Median}|}{n}$

Arrange the data in an ascending order

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| x: | 110 | 120 | 130 | 140 | 150 | 160 | 170 | 180 | 190 | 200 |

Median = $\frac{(n+1)}{2}$th item

$= \frac{(10+1)}{2}$th item

= $5.5^{th}$ item

$= \frac{(5\text{th item} + 6\text{th item})}{2}$

$= \frac{(150+160)}{2}$

= 155

| X | $\|x - \text{Median}\|$ |
|---|---|
| 110 | 45 |
| 120 | 35 |
| 130 | 25 |
| 140 | 15 |
| 150 | 5 |
| 160 | 5 |
| 170 | 15 |
| 180 | 25 |
| 190 | 35 |
| 200 | 45 |
| | $\sum\|x - A\| = 250$ |

Mean deviation = $\frac{\sum|x-\text{Median}|}{n}$

Mean deviation about median = $\frac{250}{10}$

$= 25$

Coefficient of Mean Deviation = $\frac{MD}{median}$

$= \frac{25}{155}$

$= 0.1613$

Coefficient of Mean variation = $\frac{MD}{median} \times 100$

$= \frac{25}{155} \times 100$

$= 16.13$

**Illustration 4.2.9**

Eleven people aged 18, 17, 19, 18, 17, 18, 21, 22, 18, 23, and 21 years old took part in a cricket match. Calculate mean deviation about mode.

**Solution**

The observation 18 appears four times, the observations 17 and 21 appear twice, and the others appear once. As a result, the value that appears the most is 18.

∴Mode = 18 years

| X | $\|x-\text{Mode}\|$ |
|---|---|
| 18 | 0 |
| 17 | 1 |
| 19 | 1 |
| 18 | 0 |
| 17 | 1 |
| 18 | 0 |
| 21 | 3 |
| 22 | 4 |
| 18 | 0 |
| 23 | 5 |
| 21 | 3 |
| | $\sum\|x-A\|= 18$ |

Mean deviation = $\frac{\sum|x-\text{Mode}|}{n}$

Mean deviation about mode = $\frac{18}{11}$

= 1.636

**ii. For discrete series**

**Steps,**

**For calculating the Mean Deviation.**

a. Compute the average (mean, median, or mode) required to calculate the mean deviation

b. Take the deviation of an item from the average (mean, median, or mode) i.e. $|X-A|$

c. Multiply $|x-A|$ by their frequency 'f'. Thus we get $f|x-A|$ to each observation

d. Calculate the total value of the deviation multiplied by its frequency i.e. $\sum f|x-A|$

e. Divide the total obtained by the number of observations.

$$MD = \frac{\sum f|x-A|}{N}$$

**For calculating the coefficient of mean deviation**

Coefficient of Mean Deviation = $\frac{\text{MD}}{\text{A}}$

**For calculating the coefficient of mean variation**

Coefficient of Mean variation = $\frac{\text{MD}}{\text{A}} \times 100$

**Illustration 4.2.10**

The weights of 20 students in a class are shown below.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Weight (in kg): | 49 | 53 | 54 | 55 | 66 | 68 | 70 | 80 |
| No of student: | 1 | 2 | 4 | 5 | 3 | 2 | 2 | 1 |

a. Mean deviation about mean
b. Coefficient of Mean deviation
c. Coefficient of Mean variation.

**Solution:**

| **Weight (in kg)** | **No of student (f)** | **f *x** | **\|x − Mean\|** | **F \|x − Mean\|** |
|---|---|---|---|---|
| 49 | 1 | 49 | 11 | 11 |
| 53 | 2 | 106 | 7 | 14 |
| 54 | 4 | 216 | 6 | 24 |
| 55 | 5 | 275 | 5 | 25 |
| 66 | 3 | 198 | 6 | 18 |
| 68 | 2 | 136 | 8 | 16 |
| 70 | 2 | 140 | 10 | 20 |
| 80 | 1 | 80 | 20 | 20 |
| **Total** | **N= 20** | **1200** | | $\sum f\,\|x-A\|$ **=148** |

$\bar{x} = \frac{\sum f\,x}{N}$

$= \frac{1200}{20}$

$= 60$

Mean weight = 60 kg

$$MD = \frac{\sum f\,|x - A|}{N}$$

$$MD = \frac{148}{20}$$

$$= 7.4$$

Coefficient of Mean Deviation $= \frac{MD}{A}$

$$= \frac{7.4}{60}$$

$$= 0.123$$

Coefficient of Mean variation $= \frac{MD}{A} \times 100$

$$= \frac{7.4}{60} \times 100$$

$$= 12.33$$

**Illustration 4.2.11**

Find the Mean deviation from median, Coefficient of Mean deviation, Coefficient of Mean variation for the following data

| | | | | | |
|---|---|---|---|---|---|
| x: | 10 | 11 | 12 | 13 | 14 |
| f: | 3 | 12 | 18 | 12 | 3 |

**Solution**

| X | f | Cum f | $\lvert x - \text{Median}\rvert$ | F $\lvert x - \text{Median}\rvert$ |
|---|---|---|---|---|
| 10 | 3 | 3 | 2 | 6 |
| 11 | 12 | 15 | 1 | 12 |
| 12 | 18 | 33 | 0 | 0 |
| 13 | 12 | 45 | 1 | 12 |
| 14 | 3 | 48 | 2 | 6 |
| **Total** | **n= 48** | | | $\sum f\,\lvert x - A\rvert = 36$ |

Median $= \left(\left(\frac{n+1}{2}\right)^{th}\right)$ term

$= \frac{49^{th}}{2}$ term

$= 24.5^{th}$ term

= 12

$MD = \frac{\sum f|x-A|}{N}$

$MD = \frac{36}{48}$

= 0.75

Coefficient of Mean Deviation $= \frac{MD}{A}$

$= \frac{0.75}{12}$

= 0.0625

Coefficient of Mean variation $= \frac{MD}{A} \times 100$

$= \frac{0.75}{12} \times 100$

= 6.25

**iii. For continuous series**

**Steps,**

**For calculating the Mean Deviation.**

a. Calculate the required average (mean, median, or mode) to calculate mean deviation.

b. Calculate the item's deviation from the average (mean, median, or mode), i.e., |x-A|, where x is the midpoint.

c. Multiply |x-A| by the frequency 'f' of each observation to get $f|x - A|$.

d. Calculate the deviation's total value multiplied by its frequency

   i.e $\sum f\,|x - A|$

e. Divide the total obtained from the number of observation

   $MD = \frac{\sum f\,|x-A|}{n}$

**For calculating the coefficient of mean deviation**

Coefficient of Mean Deviation $= \frac{MD}{A}$

**For calculating the coefficient of mean variation**

Coefficient of Mean variation = $\frac{\text{MD}}{\text{A}} \times 100$

**Illustration 4.2.12**

Calculate the mean deviation of the following data using the median. Find the coefficient of mean variation and the coefficient of mean deviation as well.

| Price of mango per kg: | 100-120 | 120-140 | 140-160 | 160-180 | 180-200 |
|---|---|---|---|---|---|
| Demand: | 4 | 6 | 10 | 8 | 5 |

**Solution**

Mean deviation about median = $\frac{\sum f|x - \text{Median}|}{n}$

| Class | Mid Value (x) | f | cf | $\lvert x - \text{median}\rvert$ | $f\lvert x - \text{median}\rvert$ |
|---|---|---|---|---|---|
| 100-120 | 110 | 4 | 4 | 43 | 172 |
| 120-140 | 130 | 6 | 10 | 23 | 138 |
| 140-160 | 150 | 10 | 20 | 3 | 30 |
| 160-180 | 170 | 8 | 28 | 17 | 136 |
| 180-200 | 190 | 5 | 33 | 37 | 185 |
| **Total** | | **33** | | | **661** |

N = 33

Calculation of median

$= \left(\frac{N}{2}\right)^{th}$ class

$= \left(\frac{33}{2}\right)^{th}$ class

= 16.5

The class having cumulative frequency 16.5 is 140-160

Median = $l + \frac{\frac{N}{2} - m}{f} \times c$

$= 140 + \frac{16.5 - 10}{10} \times 20$

$= 140 + \frac{130}{10}$

=140 +13

=153

$MD = \frac{661}{33}$

=20.03

Coefficient of Mean Deviation = $\frac{20.03}{153}$

= 0.130

Coefficient of Mean variation = $\frac{MD}{A} \times 100$

$= \frac{20.03}{153} \times 100$

= 13.09

**Illustration 4.2.13**

Compute the mean deviation from mode of the following data. Also find the coefficient of mean variation and the coefficient of mean deviation as well.

| | | | | | | |
|---|---|---|---|---|---|---|
| Profit per shop (Rs): | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 |
| No. of shop: | 12 | 18 | 27 | 20 | 17 | 6 |

**Solution:**

| Class | F | Mid Value (x) | \|x – mode\| | f\|x – mode\| |
|---|---|---|---|---|
| 0-10 | 12 | 5 | 20.63 | 247.56 |
| 10-20 | 18 | 15 | 10.63 | 191.34 |
| 20-30 | 27 | 25 | 0.63 | 17.01 |
| 30-40 | 20 | 35 | 9.37 | 187.4 |
| 40-50 | 17 | 45 | 19.37 | 329.29 |
| 50-60 | 6 | 55 | 29.37 | 176.22 |
| **Total** | | **100** | | **1148.82** |

Mode = l + $\frac{(f_1 - f_0)}{2f_1 - f_0 - f_2} \times c$

Mode = 20 + $\frac{(27-18)}{2*27-18-20} \times 10$

= 20 + $\frac{9}{16} \times 10$

= 20 + 5.63

= 25.63

MD = $\frac{1148.22}{100}$

= 11.48

Coefficient of Mean Deviation = $\frac{11.48}{25.63}$ = 0.45

Coefficient of Mean variation = 0.45 × 100 = 45

## 4.2.4 Standard deviation

Standard deviation 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. Karl Pearson was the first to introduce the concept of standard deviation. It is the most used methods of dispersion since it is free from some defects of other measures of dispersion.

**Advantages of standard deviation**

a) It is rigidly defined.

b) It is based on all observation.

c) Never disregards the plus or minus sign

d) It can be subjected to more mathematical analysis.

e) The changes in sampling have little effect on it.

f) It allows us to compare and contrast two or more series and determine their consistency or stability.

g) It is used in testing of hypothesis

**Disadvantages of standard deviation**

a) A layman would find it difficult to comprehend.

b) It is complex to calculate since it incorporates several mathematical models.

c) It cannot be used to compare the dispersion of two or more series of observations with different units of measurement.

**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.

$$CV = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100 \quad = \frac{\sigma}{\bar{x}} \times 100$$

**Computation of Standard deviation and Coefficient of variation**
**i.For individual series**

**Standard deviation** = $\sqrt{\frac{\sum d^2}{N}}$

Where,

d - Deviation of the item from their actual mean $(x - \bar{x})$
N - Total number of items
$d^2$ – Squares of deviation taken from actual mean

**Illustration 4.2.14**

A study looked into how many hours students spent studying before an exam. A total of eleven students were chosen. The students' responses are 8, 6, 3, 0, 5, 9, 2, 1, 3, 5, 2. Calculate the standard deviation and coefficient of variation for the number of hours that students have spent studying.

**Solution**

$$\bar{x} = \frac{\sum x}{n} = \frac{44}{11} = 4$$

| X | d (x -4) | $d^2$ |
|---|---|---|
| 8 | 4 | 16 |
| 6 | 2 | 4 |
| 3 | -1 | 1 |
| 0 | -4 | 16 |
| 5 | 1 | 1 |
| 9 | 5 | 25 |

| | | |
|---|---|---|
| 2 | -2 | 4 |
| 1 | -3 | 9 |
| 3 | -1 | 1 |
| 5 | 1 | 1 |
| 2 | -2 | 4 |
| **$\sum x = 44$** | | **$\sum d^2 = 82$** |

Standard deviation = $\sqrt{\frac{\sum d^2}{N}}$

$= \sqrt{\frac{82}{11}}$

$= \sqrt{7.4545}$

$= 2.73$

CV $= \frac{\text{Standard Deviation}}{\text{Mean}} \times 100$

$= \frac{2.73}{4} \times 100$

$= 68.26\%$

**ii. For discrete series**

**Standard deviation** = $\sqrt{\frac{\sum fx^2}{N} - \bar{x}^2}$

Where,

N - Total frequency

f – Frequency

**Illustration 4.2.15**

An arithmetic test was given to 100 kids. The following is the time in minutes required to finish the test:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Time (in minute): | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 |
| No of students: | 3 | 7 | 11 | 14 | 18 | 17 | 13 | 8 | 5 | 4 |

Calculate the standard deviation of their test completion time as well as the coefficient of variation.

| x | f | f x | $x^2$ | f $x^2$ |
|---|---|---|---|---|
| 18 | 3 | 54 | 324 | 972 |
| 19 | 7 | 133 | 361 | 2527 |
| 20 | 11 | 220 | 400 | 4400 |
| 21 | 14 | 294 | 441 | 6174 |
| 22 | 18 | 396 | 484 | 8712 |
| 23 | 17 | 391 | 529 | 8993 |
| 24 | 13 | 312 | 576 | 7488 |
| 25 | 8 | 200 | 625 | 5000 |
| 26 | 5 | 130 | 676 | 3380 |
| 27 | 4 | 108 | 729 | 2916 |
| | **N = 100** | **2238** | | **50562** |

$\bar{x} = \frac{\sum fx}{N}$

$= \frac{2238}{100}$

$= 22.38$

Standard deviation = $\sqrt{\frac{\sum fx^2}{N} - \bar{x}^2}$

$= \sqrt{\frac{50562}{100} - 22.38^2}$

$= \sqrt{505.62 - 500.8644}$

$= \sqrt{4.7556}$

$= 2.181$

CV $= \frac{\text{Standard Deviation}}{\text{Mean}} \times 100$

$= \frac{2.181}{22.38} \times 100$

$= 9.75\%$

**iii. For continuous series**

**Standard deviation =** $\sqrt{\frac{\sum fx^2}{N} - \bar{x}^2}$

Where,

N - Total frequency

f – Frequency
x – Mid value

**Illustration 4.2.16**

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.

**Solution**

| Profit | Mid value (x) | f | $x^2$ | fx | $fx^2$ |
|---|---|---|---|---|---|
| 0-10 | 5 | 8 | 25 | 40 | 200 |
| 10-20 | 15 | 12 | 225 | 180 | 2700 |
| 20-30 | 25 | 20 | 625 | 500 | 12500 |
| 30-40 | 35 | 30 | 1225 | 1050 | 36750 |
| 40-50 | 45 | 20 | 2025 | 900 | 40500 |
| 50-60 | 55 | 10 | 3025 | 550 | 30250 |
| | | | | **$\sum fx = 3220$** | **$\sum fx^2 = 122900$** |

N = 100

$\bar{x} = \frac{\sum fx}{N}$

$= \frac{3220}{100}$

$= 32.2$

Standard deviation $= \sqrt{\frac{\sum fx^2}{N} - \bar{x}^2}$

$= \sqrt{\frac{122900}{100} - 32.2^2}$

$= \sqrt{192.16}$

$= 13.86$

$$CV = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100$$

$$= \frac{13.86}{32.2} \times 100$$

$$= 43.04\ \%$$

**Illustration 4.2.17**

The results of five distinct class tests for two students, Rahul and Manu, are shown here.

| | | | | | |
|---|---|---|---|---|---|
| **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**

$$\text{Rahul's } \bar{x} = \frac{\sum x}{n}$$

$$= \frac{110}{5}$$

$$= 22$$

$$\text{Manu's } \bar{x} = \frac{\sum x}{n}$$

$$= \frac{75}{5}$$

$$= 15$$

| **Rahul** | | | **Manu** | | |
|---|---|---|---|---|---|
| **X** | **d (x -22)** | **$d^2$** | **X** | **d (x -15)** | **$d^2$** |
| 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 |
| **$\sum x = 110$** | | **$\sum d^2 = 66$** | **$\sum x = 75$** | | **$\sum d^2 = 68$** |

| Rahul | Manu |
|---|---|
| Standard deviation $= \sqrt{\frac{\Sigma d^2}{N}}$ <br> $= \sqrt{\frac{66}{5}}$ <br> $= \sqrt{13.2}$ <br> $= 3.63$ <br> $CV = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100$ <br> $= \frac{3.63}{22} \times 100$ <br> $= 16.5\%$ | Standard deviation $= \sqrt{\frac{\Sigma d^2}{N}}$ <br> $= \sqrt{\frac{68}{5}}$ <br> $= \sqrt{13.6}$ <br> $= 3.69$ <br> $CV = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100$ <br> $= \frac{3.69}{15} \times 100$ <br> $= 24.6\ \%$ |

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

$\sigma_1$ – Standard deviation of the first series

$\sigma_2$ – 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

$\bar{x}_{1.2} - \frac{N_1\bar{x}_1 + N_1\bar{x}_1}{N_1 + N_2}$

$N_1$ –Number of items of the first series

$N_2$ –Number of items of the second series

**Illustration 4.2.18**

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 items** | 50 | 40 |

**Solution**

$$\sigma_{1.2} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$$

$$\bar{x}_{1.2} = \frac{N_1\bar{x}_1 + N_1\bar{x}_1}{N_1 + N_2}$$

$$= \frac{(50 \times 63) + (40 \times 54)}{50 + 40}$$

$$= \frac{3150 + 2160}{90}$$

$$= \frac{5310}{90}$$

$$= 59$$

$d_1 = (\bar{x}_1 - \bar{x}_{1.2})$

$= (63 – 59)$

$= 4$

$d_2 = (\bar{x}_2 - \bar{x}_{1.2})$

$= (54 - 59)$

$= -5$

$$\sigma_{1.2} = \sqrt{\frac{(50 \times 8^2) + (40 \times 7^2) + (50 \times 4^2) + (40 \times -5^2)}{50 + 40}}$$

$$= \sqrt{\frac{(50 \times 64) + (40 \times 49) + (50 \times 16) + (40 \times 25)}{90}}$$

$$= \sqrt{\frac{3200 + 1960 + 800 + 1000}{90}}$$

$$= \sqrt{\frac{6960}{90}}$$

$= \sqrt{77.33}$

= 8.79

**Illustration 4.2.19**

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**

$$\sigma_{1.2} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1 d_1^2 + N_2 d_2^2}{N_1 + N_2}}$$

$$\bar{x}_{1.2} = \frac{N_1\bar{x}_1 + N_1\bar{x}_1}{N_1 + N_2}$$

$$= \frac{(550 \times 60) + (600 \times 48.5)}{550 + 600}$$

$$= \frac{33000 + 29100}{1150}$$

$$= \frac{62100}{1150}$$

= 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

**Correction in mean and standard deviation**

Sometimes, there can be errors in the data we use to calculate the average (mean) and how spread out the data is (standard deviation). These errors might only become clear after we've done the calculations. So, to get accurate results, we need to adjust the mean and standard deviation by considering the correct values for those observations. .

**Illustration 4.2.20**

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 $\sum x = \overline{x} \times n$

$= 5 \times 11$

$= 55$

Correct $\sum x$ = Incorrect $\sum x$ – wrong item + correct item

= 55 – 13 + 2

= 44

Correct $\overline{x} = \frac{44}{11}$

= 4

Calculation of the correct Standard Deviation

$$\sigma = \sqrt{\frac{\sum x^2}{N} - (\overline{x})^2}$$

$3.67 = \sqrt{\frac{\sum x^2}{11} - (5)^2}$

Squaring both sides

$3.67^2 = \frac{\sum x^2}{11} - 25$

$13.4689 + 25 = \frac{\sum x^2}{11}$

$38.4689 = \frac{\sum x^2}{11}$

$\sum x^2 = 38.4689 \times 11$

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 4.2.21**

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 $\sum x = \bar{x} \times n$

$= 40 \times 200$

$= 8000$

Correct $\sum x$ = Incorrect $\sum x$ – wrong item + correct item

= 8000 – (34 + 53) + (43 + 35)

= 8000 – 87 +78

= 7991

Correct $\bar{x} = \frac{7991}{200}$

= 39.955

Calculation of the correct Standard Deviation

$$\sigma = \sqrt{\frac{\sum x^2}{N} - (\bar{x})^2}$$

$$15 = \sqrt{\frac{\sum x^2}{200} - (40)^2}$$

Squaring both sides

$$15^2 = \frac{\sum x^2}{200} - 1600$$

$$1600 + 225 = \frac{\sum x^2}{200}$$

$$1825 = \frac{\sum x^2}{200}$$

$\sum x^2 = 1825 \times 200$

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

## Recap

- The range - difference between the largest and the least numbers in the set.
- Quartile Deviation - half of the difference between the upper and lower quartile.
- Inter-quartile range - The difference between upper quartile and lower
- The mean deviation - statistical measure used to calculate the average deviation from the average value of the series.
- The standard deviation - statistic that calculates the square root of the variance and measures the dispersion of a dataset relative to its mean.
- The standard deviation - the value of standard deviation cannot be negative.
- The coefficient of variation (CV) - ratio of the standard deviation to the mean.

## Objective Questions

1. What is the name of the average calculated from a set of observations in which a single number is picked based on its position to represent the whole set?
2. What average is obtained when an inter quartile range is divided by 2?
3. What are the different types of averages that are used to calculate the mean deviation?
4. Which type of average is required to calculate standard deviation?
5. What term is used to denote the positive square root of the mean of the squares of deviation from the arithmetic mean?
6. What statistical tool is used to compare the consistency of 2 variables?
7. What term is used to denote the difference between the highest and lowest value in a series?
8. Who introduced standard deviation?

## Answers

1. Positional average
2. Semi-interquartile range
3. Mean, Median and Mode.
4. Mean
5. Standard deviation
6. Coefficient of variation
7. Range
8. Karl Pearson

## Self-Assessment Questions

1. Describe the concept of "Variance" as a measure of dispersion. Discuss its calculation and the relationship between variance and standard deviation.
2. What is the "Coefficient of Variation (CV)"? How is it calculated, and how does it help in comparing the relative variability of two or more data sets?
3. Differentiate between "Range" and "IQR" as measures of dispersion. What are the advantages and limitations of each?
4. Discuss the importance of "Standard Deviation" in statistics. Explain its formula and why it is widely used in analysing data variability.
5. Explain the concept of "Coefficient of Quartile Deviation (QD)." How is it calculated, and what insights does it provide regarding data distribution?
6. Describe the role of measures of dispersion in data analysis and decision making. Provide examples of situations where understanding data spread is crucial

| 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**

3. Marks obtained by ten students in a class test is given below.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Marks: | 20 | 25 | 40 | 30 | 35 | 45 | 70 | 65 | 60 | 80 |

Find Mean Deviation from mean and coefficient of Mean Deviation.

**Ans. M.D = 17.4, Coefficient of M.D = 0.37**

4. The following table shows the number of books read by students in a B.com class consisting of 28 students, in a month.

| | | | | | |
|---|---|---|---|---|---|
| No of Books: | 0 | 1 | 2 | 3 | 4 |
| No of students: | 2 | 6 | 12 | 5 | 3 |

Calculate mean deviation about mode of number of books read.

**Ans. 0.75**

5. Calculate mean deviation about mean of the number of telephone calls received at an exchange:

| No of calls: | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|---|---|
| Frequency: | 4 | 6 | 10 | 20 | 10 | 6 | 4 |

**Ans. 11.33**

6. Find standard deviation and coefficient of variation from the following data.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| X: | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

**Ans. S.D= 28.72, CV = 52.21%**

7. The score of two batters Saju and Raju in 10 innings during a certain season are as follows

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Saju | 25 | 50 | 45 | 30 | 70 | 42 | 36 | 48 | 34 | 60 |
| Raju | 10 | 70 | 50 | 20 | 95 | 55 | 42 | 60 | 48 | 80 |

Find which of the two batters, Saju or Raju is more consistent in scoring? Also state who is better run getter?

**Ans. Saju - $\bar{x}$ = 44, SD = 13.08, CV = 29.37% Raju - $\bar{x}$ = 53, SD =24.35, CV = 45.94. Saju is more consistent in Scoring as his CV is low. Raju is a better run getter as his arithmetic mean is higher than that of Saju.**

8. A factory produces two types of batteries X and Y. in an experiment relating to their life the following result were obtained.

| Length of life (Hrs) | Number of battery X | Number of Battery Y |
|---|---|---|
| 500-700 | 5 | 4 |
| 700-900 | 11 | 30 |
| 900-1100 | 26 | 12 |
| 1100-1300 | 10 | 8 |
| 1300-1500 | 8 | 6 |

a) Which battery X or Y has more average life?

b) Which is more consistent?

**Ans. (a) Battery X has more average life.**

**(b) Battery X is more consistent.**

9. A student is obtained the mean and Standard Deviation of 100 observation as 40 and 5.1 respectively. It was later discovered that the value of item 40 was misread as 50. Calculate the correct mean and standard deviation.

**Ans. Mean = 39.9, S.D = 4.3**

## Suggested Readings


1. C.R.Kothari (2013). *Quantitative Techniques*. Vikas Publishing House.
2. S.P.Guptha-(2012). *Statistical Methods*. Sultan Chand &Sons,New Delhi
3. C. B Gupta & Vijay Gupta (2004). *An Introduction to Statistical Methods*. Vikas Publishing
4. Goel & Goel (2014). *Mathematics & Statistics*. Taxmann Allied Services Pvt Ltd
5. Neiswanger, William A. *Elementary Statistical Methods*, The Macmillian Co., New York.
6. Richard I. Levin and David S. Rubin. *Statistics for Management*. Seventh Edition. Pearson—Prentice Hall of India, New Delhi.
7. Riggleman and Frisbee. *Business Statistics*, McGraw Hill Book Co., New York.

5

BLOCK

# Correlation and Regression Analysis

# Unit - 1
# Correlation

## Learning Outcomes

After going through the unit "Correlation" the learner will be able to:

- ✓ learn the concepts of correlation
- ✓ acquire knowledge of the different types of correlation
- ✓ get an idea about the nature of the relationship between two variables
- ✓ be aware of the importance and uses of correlation in the day-to-day life

## Prerequisites

In a town close to Munnar, Ajay and Leela ran two small businesses that did very well. Ajay sold freshly brewed spiced chai to both locals and visitors, and Leela sold hand-knit woollen scarves to people who came to the cold hill station. As the months went by, they saw that Ajay's chai sales and Leela's scarf sales both went up when the temperature dropped. It was their teacher who told them that this link was an example of association.

It was told by the teacher that correlation shows how two variables are linked. In this case, the cold weather made more people want both chai and scarves, so there was a good relationship between their sales. But the teacher also made one important point clear: a link does not mean a cause. During the Kerala rainfall, for example, both the sale of hot soups and raincoats go up, but neither one causes the other. They are both caused by the rainy season. Once Ajay and Leela knew this, they started to prepare their businesses for the colder months by stocking up and making better plans, which helped them make more money.

## Keywords

Positive and negative correlation, Linear and non-linear correlation, Partial and Multiple Correlation

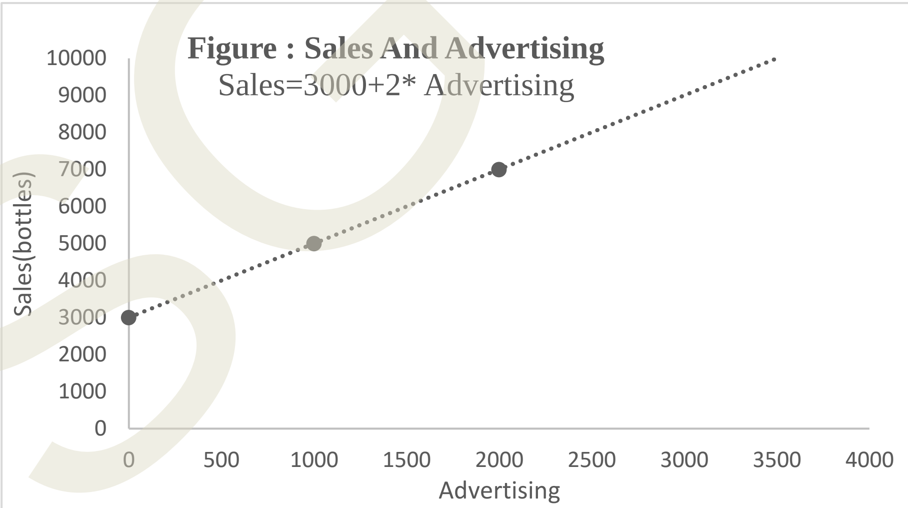
## Discussion

### 5.1.1 Introduction

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. 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 of a 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.

Consider the figure,




In the figure 'advertising' is an independent variable as it can accept any value but sales depend upon advertising and hence 'Sales' is a dependent variable. In this Bivariate analysis 'Sales' is related to '3000 + 2* Advertising'. We can forecast the sale value if advertisement cost is known. This measure of association between advertising and sales is called Correlation.

## 5.1.2 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 increase in one variable corresponds to an increase or decrease in the other variable, and similarly, a decrease 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.

## 5.1.3 Types of Correlation

Correlation is classified into the following types.

- Positive and negative Correlation
- Linear and non-linear Correlation
- Simple, partial and multiple Correlation

### Positive and Negative correlation

The direction of correlation, whether it is positive (indicating a direct relationship) or negative (indicating an inverse relationship), hinges on how the variables change. When both variables move in the same direction, meaning that as one increases, the other, on average, also increases, or as one decreases, the other, on average, decreases, we classify this as a positive correlation. For instance, the more time you spend running on a treadmill, the more calories you burn, illustrating a positive correlation. The relationship between income and expenditure is another example of a positive correlation.

Conversely, when the variables vary in opposite directions, meaning that as one increases, the other decreases, or vice versa, we refer to this as a negative correlation. For example, the correlation between the price of a commodity and its demand is a negative correlation. Similarly, the relationship between volume and pressure in a gas follows a negative correlation pattern.

### Linear and Non-linear correlation

Consider the values of two variables.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| X | 2 | 4 | 6 | 8 | 10 | 12 | 14 |
| Y | 4 | 8 | 12 | 16 | 20 | 24 | 28 |

It is observed that ratio of change between the two variables remains constant. If the ratio of change between the two variables is a constant value, then the correlation between the variables is called linear correlation. If we plot these points on a graph, we get a straight line.

However, if the amount of change in one variable does not bear a constant ratio with the amount of change in the other variable, the relation is called curvi-linear (non-linear correlation) and the resultant graph will be a curve.

## Simple, Partial, and Multiple correlation

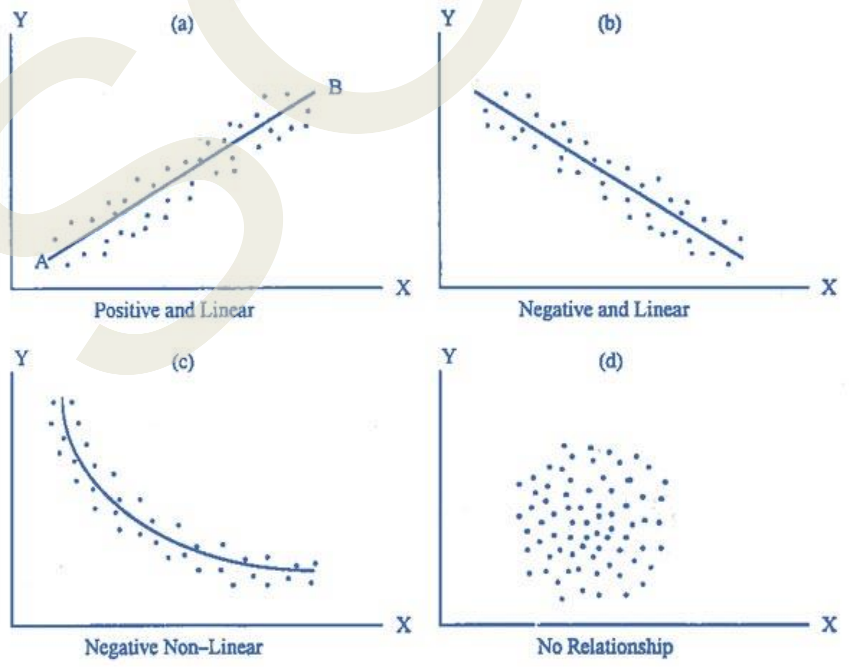
This distinction is based on the number of variables under study. When examining only two variables, it is referred to as simple correlation. For instance, if we study the relationship between the yield of paddy and the fertilizer used, it is a simple correlation. When three or more variables are studied, it is either multiple or partial correlation. In multiple correlation, three or more variables are studied simultaneously. For example, when we study the relationship between the yield of rice per acre, with the amount of rainfall, and the quantity of fertilizers used, it is an example of multiple correlation.

On the other hand, in partial correlation, we study more than two variables, but consider only two variables that influence each other, while keeping the effect of other influencing variables constant. In the study of production of rice, if we consider yield and rainfall only, maintaining the daily temperature as constant, it illustrates an example of partial correlation.

## No correlation

If the variables do not have a relationship with each other, then there is no correlation.

Diagrammatic Representation



### 5.1.4 Uses of correlation

i. **Financial:** Correlation is used in finance to analyse the relationship between different financial assets.

ii. **Economics:** Correlation analysis helps in analysing the economic behaviour. For example, wages and the level of inflation, savings and the rate of tax collected etc. By understanding these connections, economists can propose strategies to enhance the prevailing situation.

iii. **Social sciences:** The importance of analysing and understanding the relationship between two or more variables is increasing day by day with the developments in the field of social sciences. For example, to study the relationship between smoking and lung cancer, number of accidents and the sale of vehicles etc.

iv. **Climate science:** Climate scientists use correlation to analyse relationship between different climate variables such as temperature and atmospheric levels to understand how they affect each other.

v. **Education:** Correlation can be used in educational research to study relationship between variables like teaching methods and academic performance.

### 5.1.5 Importance of correlation

The following are the importance of correlation:

i. Understanding relationships:

Correlation serves as a tool to gauge relationships between variables. Imagine you are in charge of human resources at a company. You want to assess whether the level of training provided to employees is linked to their job performance. Correlation can help you understand this connection, much like evaluating the relationship between studying for an exam and getting a good grade.

ii. Predicting trends:

Correlation empowers commerce professionals to foresee trends. For instance, in the stock market, understanding the correlation between two stocks enables traders to make more informed predictions.

iii. Identifying key factors:

In the complex world of commerce data, correlation helps identify which variables are most crucial. This is analogous to picking the essential ingredients for a recipe. For accountants, it streamlines financial analysis by spotlighting interconnected metrics.

iv. Quality control:

Consistency is paramount in commerce. Correlation ensures that changes in one part of a process reflect accurately in another, making it indispensable for quality control in manufacturing and services.

v. Risk management:

Correlation aids in managing risks, much like playing cards strategically. Understanding how different variables correlate helps investors diversify portfolios and reduce potential losses.

vi. Optimising resource allocation:

In commerce, resources such as advertising budgets must be allocated judiciously. Correlation helps identify which strategies yield the best results, enabling businesses to optimise their spending.

vii. Enhanced decision making:

Correlation assists in data-driven decision making. For instance, it reveals whether online or television advertising has a greater impact on sales, facilitating informed choices.

viii. Refining models:

In economic modeling, correlation is pivotal for constructing accurate models. It is similar to selecting the right ingredients and proportions for a recipe, ensuring the model mirrors real-world dynamics faithfully.

## 5.1.6 Limitations of correlation

**The following are the limitations of correlation:**

i. Correlation measures only linear relationships, it may miss significant non-linear associations between variables.
ii. Extreme values can disproportionately affect correlation, leading to inaccurate results.
iii. Correlation captures only strength and direction, not the full nature of relationships.
iv. Small samples might not represent the population accurately, affecting correlation results.
v. Uncontrolled or unaccounted variables might influence the relationship between the correlated variables, leading to misleading interpretations.

## Recap

- If two quantities vary in such a way that movements in one are accompanied by movements in the other, then these quantities are said to be correlated.
- Positive correlation- if both the variables are varying in the same direction,
- Negative correlation- if one variable is increasing, the other is decreasing, or vice versa.
- Linear correlation- if the ratio of change between the two variables is constant.
- Non-linear correlation- if the amount of change in one variable does not bear a constant ratio of the amount of change in the other.
- Simple correlation- When only two variables are studied.
- Multiple correlation- When three or more variables are studied.
- Partial correlation- If we study only two variables and eliminates some other factors or variable.
- No correlation- the variables do not have a relationship with each other.

## Objective Questions

1. What is the measure of association between two variables called?
2. If two variables move in opposite direction, which type of correlation is it?
3. Which type of correlation exist, if we consider more than two variables in correlation study?
4. If the graph of the data is a straight line, which type of correlation is revealed between the elements in the data?
5. If the ratio of change between the two variables is not constant, which type of correlation exist?
6. Name any four types of correlation.
7. What does a positive correlation between two variables indicate?
8. What is the primary goal for conducting correlation analysis?

## Answers

1. Correlation
2. Negative correlation
3. Multiple correlation
4. Linear correlation
5. Non-Linear correlation
6. Linear, non-linear, simple, partial, multiple
7. Direct relationship
8. Measure the strength and direction of the relationship

## Self-Assessment Questions

1. What is linear and nonlinear correlation?
2. Bifurcate simple, partial and multiple correlation.
3. Explain the different types of correlation?
4. What are the uses of correlation?
5. Write a short note on the limitations of correlation.

## Assignments

1. Give an example of a positive correlation that you've noticed in your daily experiences?
2. What is correlation, and why is it important in everyday life?
3. In the world of commerce, do you think there's a correlation between customer satisfaction and repeat business? Can you explain why or why not?

## Suggested Readings


1. S.P.Guptha-(2012). *Statistical Methods*. Sultan Chand &Sons,New Delhi
2. Goel & Goel (2014). *Mathematics & Statistics*. Taxmann Allied Services Pvt Ltd
3. Secrist Horace. *An Introduction to Statistical Methods*, The Macmillan Co., New York.
4. Simpson and Kafka. *Basic Statistics*, Oxford and I.B.H. Publishing Co., Calcutta.

# Unit - 2
# Measure of Correlation

## Learning Outcomes

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

- ✓ get an idea on the different measures of correlation
- ✓ compute rank correlation coefficient when ranking rather than actual values for variables are known
- ✓ analyse the degree and direction of the relation
- ✓ appreciates some practical applications of correlation

## Prerequisites

Correlation helps us understand how two variables are related. For instance, in a city like Jaipur, imagine Ramesh, who sells umbrellas, and Priya, who runs a raincoat shop. They notice that on rainy days, their sales increase simultaneously. This is an example of a positive correlation, where both variables increase together. Alternatively, if Ramesh observes that on sunny days, his umbrella sales drop while Priya's raincoat sales remain steady, this could indicate no significant relationship between the two variables.

To measure such relationships, different methods of correlation are used based on the type of data. When the data shows a clear, linear pattern - like daily sales numbers of umbrellas and raincoats - a method designed for continuous data can measure the strength and direction of this relationship. For example, if both sales rise in a proportional manner, a straightforward calculation can determine how closely they are connected.

In some cases, data may not follow a straight pattern. Let's say Ramesh and Priya conduct a customer survey, asking people to rank their preferences for umbrellas and raincoats. Here, ranked or ordinal data requires a different method to assess the correlation. This approach works best when the relationship is not linear but still provides meaningful insights, like how preferences for one product affect the other.

Choosing the right method to measure correlation allows people like Ramesh and Priya to analyze patterns effectively. It helps them plan their stock, predict demand, and improve sales strategies. Understanding correlation is not just limited to businesses - it's a key tool for making data-driven decisions across various fields, from education to healthcare and beyond.

## Keywords

Scatter diagram, Coefficient of correlation, Karl Pearson's coefficient of correlation, Rank correlation.

## Discussion

### 5.2.1 Methods of Studying Correlation

As mentioned in the previous unit, the statistical tool used to explore the relationship between two or more variables are known as correlation. To determine the linearity and non- linearity among the variables, and the extent to which they are correlated, various methods are used to define and measure the correlation among the variables. The various methods of studying correlation coefficient are:

**Scatter Diagram Method**

This method is the simplest and the easiest method for studying correlation. The two variables X and Y are plotted on a two-dimensional graph. One variable is represented along X axis and the other variable along Y axis. The graph thus plotted will represent the relationship between the variables. Thus, this study of relationship between two variables based on the graphical representation is called scatter diagram method. This method is unscientific and not popular nowadays due to the availability of more advanced methods of studying correlation which we will learn as Karl Pearson's method and Spearman's rank correlation methods. This method is influenced by the personal judgement of the investigator and we use the advanced methods for studying correlation between the variables.

**Coefficient of correlation**

Degree of relationship between two variables is called coefficient of correlation. It is an algebraic method of measuring correlation. Coefficient of correlation is denoted by the symbol r and the value of r lies between −1 and +1.

i.e., $-1 \leq r \leq 1$

## Properties of Correlation coefficient

The following are the properties of correlation coefficient:

i. Coefficient of correlation lies between −1 and +1.

ii. When $r$ lies between 0 and 1, the correlation is positive, when $r$ lies between -1 and 0, the correlation is negative. If r = 0 there is no correlation.

iii. If $r = +1$ it is perfect positive correlation. If $r = -1$ it is called perfect negative correlation.

iv. It is a pure number that lies between -1 and +1 and has no units.

v. Independent variables are uncorrelated but correlation coefficient of X and Y is same as Correlation coefficient of Y and X

vi. Correlation coefficient does not change with reference to change of origin.

**Karl Pearson's Coefficient of Correlation**

This method is the most widely used method for measuring correlation. It is popularly known as Pearson 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 statistical 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.

$$Cov\ (x, y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{n}$$

$\sigma(x) = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$ is the standard deviation of $x$.

$\sigma(y) = \sqrt{\frac{\sum(y-\bar{y})^2}{n}}$ is the standard deviation of $y$.

$\bar{x} = \frac{\sum x}{n}$, $\bar{y} = \frac{\sum y}{n}$,

So, $r(x,y) = \frac{\frac{\sum(x-\bar{x})(y-\bar{y})}{n}}{\sqrt{\frac{\sum(x-\bar{x})^2}{n}}\sqrt{\frac{\sum(y-\bar{y})^2}{n}}} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}$

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 5.2.1**

Find the Karl Pearson's coefficient between x and y for the following data

$$n = 10, \sum x = 35, \sum x^2 = 203, \sum y = 28, \sum y^2 = 140, \sum xy = 168$$

**Solution**

$$r(x,y) = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

$$= \frac{10 \times 168 - 35 \times 28}{\sqrt{10 \times 203 - 35^2}\sqrt{10 \times 140 - 28^2}}$$

$$= \frac{1680 - 980}{\sqrt{805}\sqrt{616}}$$

$= 0.99$

**Illustration 5.2.2**

Find Karl Pearson's correlation coefficient between $x$ and $y$ for the following data,

$n = 15$, $Cov(x,y) = 8.13$, $\sigma(x) = 3.01$, $\sigma(y) = 3.03$

**Solution**

$$r = \frac{Cov(x,y)}{\sigma(x)\sigma(y)} = \frac{8.13}{3.01\ x\ 3.03} = 0.89$$

**Illustration 5.2.3**

Find Karl Pearson's correlation coefficient between x and y for the following data,

$$n = 1000,\ \ \sigma(x) = 4.5, \sigma(y) = 3.6, \sum(x - \bar{x}).(y - \bar{y}) = 4800$$

**Solution**

$$r(x,y) = \frac{Cov\ (x,y)}{\sigma(x)\sigma(y)}$$

$$.\ r = \frac{\frac{\sum(x-\bar{x})(y-\bar{y})}{n}}{\sigma(x)\sigma(y)}$$

$$= \frac{\frac{4800}{1000}}{4.5 \times 3.6}$$

$$= \frac{4.8}{16.2}$$

$$= 0.296$$

**Illustration 5.2.4**

Find Product Moment method of Correlation Coefficient between $x$ and $y$ for the following data,

$$n = 20, \Sigma(x - \overline{\boldsymbol{x}})^2 = 136, \Sigma(y - \overline{\boldsymbol{y}})^2 = 138,\ \ \sum(x - \bar{x}).(y - \bar{y}) = 122$$

**Solution**

$$r = \frac{\sum(x-\bar{x})\ (y-\bar{y})}{\sqrt{\sum(x-\overline{x})^2}\ \sqrt{\sum(y-\overline{y})^2}}$$

$$= \frac{122}{\sqrt{136}\,\sqrt{138}}$$

$$= \frac{122}{11.66 \times 11.75}$$

$$= 0.89$$

**Illustration 5.2.5**

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**

$$\bar{x} = \frac{\sum x}{n} = \frac{15}{5} = 3$$

$$\bar{y} = \frac{\sum y}{n} = \frac{15}{5} = 3$$

| $x$ | $y$ | $x-3$ | $y-3$ | $(x-3)^2$ | $(y-3)^2$ | $(x-3)(y-3)$ |
|---|---|---|---|---|---|---|
| 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** |

$$r = \frac{\sum(x-\bar{x})\,(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\,\sqrt{\sum(y-\bar{y})^2}} = \frac{6}{\sqrt{10}\,\sqrt{10}} = \frac{6}{10} = 0.6$$

**Illustration 5.2.6**

Calculate the coefficient of correlation for the following table by Karl Pearson's coefficient of correlation.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| X: | 70 | 69 | 68 | 67 | 66 | 65 | 64 |
| Y: | 72 | 68 | 70 | 68 | 65 | 67 | 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** |

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}$$

$$= \frac{25}{\sqrt{28}\sqrt{34}}$$

$$= \frac{25}{30.85}$$

$$= 0.81$$

**Illustration 5.2.7**

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**

$$\bar{x} = \frac{\sum x}{n} = \frac{108}{9} = 12$$

$$\bar{y} = \frac{\sum y}{n} = \frac{45}{9} = 5$$

| $x$ | $y$ | $x-12$ | $y-5$ | $(x-12)^2$ | $(y-5)^2$ | $(x-12)(y-5)$ |
|---|---|---|---|---|---|---|
| 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** |

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}$$

$$= \frac{57}{\sqrt{60}\sqrt{60}}$$

$$= \frac{57}{60}$$

$$= 0.95$$

**Illustration 5.3.8**

Calculate the Pearson coefficient of correlation from the following data

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| X: | 77 | 60 | 30 | 53 | 14 | 35 | 90 | 25 | 56 | 60 |
| Y: | 35 | 38 | 60 | 40 | 50 | 40 | 35 | 56 | 34 | 42 |

**Solution**

$$\bar{x} = \frac{\sum x}{n} = \frac{500}{10} = 50$$

$$\bar{y} = \frac{\sum y}{n} = \frac{430}{10} = 43$$

| $x$ | $y$ | $x-50$ | $y-43$ | $(x-50)^2$ | $(y-43)^2$ | $(x-50)(y-43)$ |
|---|---|---|---|---|---|---|
| 77 | 35 | 27 | -8 | 729 | 64 | -216 |
| 60 | 38 | 10 | -5 | 100 | 25 | -50 |
| 30 | 60 | -20 | 17 | 400 | 289 | -340 |
| 53 | 40 | 3 | -3 | 9 | 9 | -9 |
| 14 | 50 | -36 | 7 | 1296 | 49 | -252 |
| 35 | 40 | -15 | -3 | 225 | 9 | 45 |
| 90 | 35 | 40 | -8 | 1600 | 64 | -320 |
| 25 | 56 | -25 | 13 | 625 | 169 | -325 |
| 56 | 34 | 6 | -9 | 36 | 81 | -54 |
| 60 | 42 | 10 | -1 | 100 | 1 | -10 |
| **500** | **430** | | | **5120** | **760** | **-1531** |

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}} = \frac{-1531}{\sqrt{5120}\sqrt{760}} = \frac{-1531}{71.55 \times 27.57} = -0.77$$

## 5.2.2 Probable Error

The probable error of the coefficient of correlation is a statistical measure that provides an estimate of the likely amount of error in the calculated correlation coefficient. Therefore, with the help of probable error it is possible to determine the reliability of the value of the coefficient in so far as it depends on the condition of random sampling.

The probable error of the coefficient of correlation is obtained as follows;

$$P.E = 0.6745\ \frac{1 - r^2}{\sqrt{N}}$$

Where,

P. E = the notation for Probable Error.

0.6745 = The Standard Error which is a constant value derived from statistical table.

r = The Pearson coefficient of correlation.

N = Number of paired items.

**The method of interpretation**

The interpretation of the computed value of Probable error is explained below:

i. If the value of "r" is less than the probable error, there is no evidence of correlation, that is the value of "r" is not at all significant.

ii. If the value of "r" is more than six times the probable error, the existence of correlation is practically certain, that is, the value of "r" is significant.

iii. By adding and subtracting the value of probable error from the coefficient of correlation, we get respectively the upper and lower limits within which coefficient of correlation in the population can be expected to lie, symbolically;

$$\rho = r \pm P.E$$

ρ is read as (rho) denotes correlation in the population

## 5.2.3 Degree of Correlation

This is used to interpret the computed value of Karl Pearson's correlation coefficient. When we learn probable error, there is a procedure to interpret the value of correlation. But this method of interpretation is based on some general guidelines. The guidelines used in this interpretation are the following:

- Perfect Correlation: If the computed correlation coefficient is very close to either positive or negative one, it is an indication of perfect correlation. If the computed value is positive "one", it is perfect positive correlation. If the computed correlation value is negative "one", it is a perfect negative correlation. In the case of perfect positive correlation, if one variable is increasing, the other variable, on an average, is also increasing. In the case of perfect negative correlation, as one variable increases the other variable, on an average, tends to decrease.

- High degree of correlation: if the computed value of correlation coefficient lies between plus or negative 0.50 to plus or negative one, it is considered as a high degree of correlation.

- Moderate degree of correlation: If the computed value of correlation is between plus or negative, 0.30 to plus or negative 0.49, it is considered as moderate degree of correlation.

- Low degree of correlation: If the computed value of correlation is below plus or minus 0.29, then it is considered as low degree of correlation.

- Zero correlation: If the computed value of correlation is zero, then it is considered as zero correlation or there is no relationship between the variables.

## 5.2.4 Corrected Correlation Coefficient

The Corrected Correlation Coefficient, often referred to as the Attenuation-Corrected Correlation Coefficient, is a statistical measure used to assess the relationship between two variables, while accounting for measurement errors or unreliability in the data. When the observed correction coefficient is calculated from data that contains measurement errors, it tends to underestimate the true strength of the underlying relationship between the variables.

The idea behind the corrected correlation coefficient is to adjust the observed correlation coefficient to account for the effect of measurement errors. This correlation helps to provide a more accurate estimate of the actual correlation between the variables, assuming that the measurement errors are known or can be estimated.

**Illustration 5.2.9**

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** |

$$\bar{x} = \frac{\sum x}{n} = \frac{610}{10} = 61$$

$$\bar{y} = \frac{\sum y}{n} = \frac{640}{10} = 64$$

$$r = \frac{\sum(x-\bar{x})\,(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\,\sqrt{\sum(y-\bar{y})^2}} = \frac{3535}{\sqrt{3490}\,\sqrt{4390}} = \frac{3535}{59.08 * 66.26} = 0.903.$$

b. Probable error of correlation coefficient is calculated using the equation;

$$P.E = 0.6745\,\frac{1-r^2}{\sqrt{N}}$$

Substituting the values in the equation, we get

$$P.E = 0.6745\,\frac{1-(0.903)^2}{\sqrt{10}}$$

$= 0.0393$

Here the value of $r = 0.903$ and $P.E = 0.0393$

Since the value of "r" is more than P.E, we have to check whether the value of "r" is significant or not, that means the "r" value should be more than six times the probable error, to conclude that the value of "r" is significant. We have

$r = 0.9031, \quad P.E = 6x0.0393 = 0.2358$

$Here\ r > 6 \times P.E$

So, we can conclude that the value of "r" is highly significant. This means that higher the marks of a candidate in Mathematics, higher the marks in Statistics also and lower the marks of a candidate in Mathematics, lower is his score in Statistics also.

**c.** Limits for population correlation coefficient is given by the formula

$r \pm$ P.E

That means we have to add and subtract the value of "r" to the probable error value to find out the limits of population. So, we get

$0.903 \pm 0.0393 = 0.8637\ to\ 0.9423$

This implies that if we take another sample of ten items from the same population, then its correlation coefficient is expected to lie between 0.8637 and 0. 9423.

**Illustration 5.2.10**

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 and interpret the result by finding the Probable Error.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 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** |

$$\bar{x} = \frac{\sum x}{n} = \frac{153}{9} = 17$$

$$\bar{y} = \frac{\sum y}{n} = \frac{378}{9} = 42$$

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\overline{x})^2}\sqrt{\sum(y-\overline{y})^2}} = \frac{88}{\sqrt{60}\sqrt{284}}$$

$$= \frac{88}{7.75 \times 16.85}$$

$$= \frac{88}{130.6}$$

$= 0.67$

$$P.E = 0.6745 \frac{1-0.67^2}{\sqrt{9}}$$

Substituting the values in the equation, we get

$$P.E = 0.6745 \frac{1-(0.67)^2}{3}$$

$$= \frac{0.6745 \times 0.5511}{3} = 0.123$$

Here $r > P.E$, so we have to check whether $r > 6 \times P.E = 6 \times 0.123 = 0.736$

But, $r < 6 \times P.E$, hence the value of r is not significant.

**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. A Spearman correlation coefficient is also referred to as Spearman rank correlation or Spearman's rho. It is typically denoted either with the Greek letter rho (ρ).

$$\rho = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

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:

i. When actual ranks are given,
ii. 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 $\rho = 1 - \frac{6\sum D^2}{n(n^2-1)}$

**Illustration 5.2.11**

Following are the marks obtained by 8 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** |

$$\rho = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

$$= 1 - \frac{6*10}{8(8^2-1)}$$

$$= 1 - \frac{60}{504}$$

$$= 1 - 0.12$$

$$= 0.88$$

**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 5.2.12**

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** |

$$\rho = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

$$= 1 - \frac{6*6}{7(7^2-1)}$$

$$= 1 - \frac{36}{336}$$

$$= 1 - 0.107$$

$$= 0.893$$

**Illustration 5.2.13**

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** |

$$\rho = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

$$= 1 - \frac{6*40}{9(9^2-1)}$$

$$= 1 - \frac{240}{720}$$

$$= 0.6667$$

**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 $1 - \frac{m^3-m}{12}$ 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:

$$\rho = 1 - \frac{6[(\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \ldots\ldots]}{n(n^2 - 1)}$$

**Illustration 5.2.14**

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 is the highest value, so it is given 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 $\frac{7+8+9}{3} = 8^{th}$ rank. So, rank 8 will be assigned to all the values of 16. The last value 9 gets the 10$^{th}$ 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;

$$\rho = 1 - \frac{6[(\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \ \dots\dots]}{n\,(n^2 - 1)} + \cdots$$

$$= 1 - \frac{6[(52 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)]}{10^3 - 10}$$

$$= 1 - \frac{6(52+2+0.5+0.5)}{990}$$

$$= 1 - \frac{6 \times 55}{990}$$

= 0.667

**Illustration 5.2.15**

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 | 0.5 | 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 | 0.5 | 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 01 - $\frac{(2^3-2)}{12}$ is added.

Similarly for Y series value 40 is repeated thrice, so the correction factor 01 - $\frac{(3^3-3)}{12}$ is added.

The rank correlation coefficient can be calculated using the equation;

$$\rho = 1 - \frac{6[(\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \ldots\ldots]}{n\,(n^2 - 1)} + \cdots$$

$$= 1 - \frac{6[(81.5 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(3^3 - 3)]}{8^3 - 8}$$

$$= 1 - \frac{6[(81.5 + 0.5 + 2]}{8^3 - 8}$$

$$= 1 - \frac{6 \times 84}{504}$$

$$= 1 - 1 = 0$$

## Concurrent Deviation method

In this method, only the directions of deviations are taken and magnitudes are ignored. If the deviations of the two variables are concurrent then they move in the same direction, otherwise in the opposite direction.

The formula for finding coefficient of concurrent deviation is $\pm\sqrt{\pm\left(\frac{(2C-N)}{N}\right)}$

Where,

c the number of concurrent deviations

N is the number of pairs of observations

**Illustration 5.2.16**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| X: | 13 | 18 | 23 | 8 | 21 | 25 | 28 | 10 | 22 |
| Y: | 23 | 11 | 17 | 3 | 23 | 18 | 8 | 23 | 20 |

**Solution**

| X | Y | Direction of change in X Dx | Direction of change in Y Dy | DxDy |
|---|---|---|---|---|
| 13 | 23 | ------- | ----- | ------ |
| 18 | 11 | + | - | - |
| 23 | 17 | + | + | + |
| 8 | 3 | - | - | + |
| 21 | 23 | + | + | + |
| 25 | 18 | + | - | - |
| 28 | 8 | + | - | - |
| 10 | 23 | - | + | - |
| 22 | 20 | + | - | - |

C = Number of + signs in DxDy column =3

N =8 (first pair of observation is not compared)

$$r = \pm\sqrt{\pm\frac{2C-N}{N}}$$

$$= \pm\sqrt{\pm\frac{2*3-8}{8}}$$

$$= \pm\sqrt{\pm\frac{-2}{8}}$$

$$= \pm\sqrt{\pm(-0.25)}$$

= -0.5

Note:

a) Increase in the value is denoted by + sign and decrease by – sign
b) r is +ve when 2C>N and r is -ve when 2C<N

## Recap

- Coefficient of correlation- Degree of relationship between two variables.
- Karl Pearson's coefficient of correlation- Most widely used method for measuring correlation.
- Probable error- Statistical measure that provides an estimate of the likely amount of error in the calculated correlation coefficient.
- Degree of correlation-- Used to interpret the computed value of Karl Pearson's correlation coefficient.
- Rank correlation- Correlation coefficient obtained from the ranks.
- Concurrent deviation method- Only the directions of deviations are taken and magnitudes are ignored.

## Objective Questions

1. What is the purpose of measuring correlation between variables?
2. What does Karl Person correlation coefficient measure?
3. Explain the interpretation of a correlation coefficient value of -0.7.
4. What is perfect positive correlation?
5. What is the range of values for the Karl Pearson correlation coefficient?
6. What does it indicate that the correlation coefficient is -1?
7. List any two methods to study correlation.
8. What does it imply, if the computed value of correlation coefficient lies between plus or negative 0.50 to plus or negative one?

## Answers

1. To quantify the strength and direction of a linear relationship between two variables.
2. Karl Person correlation coefficient measures the linear relationship between two continuous variables.
3. There is a strong negative linear relationship between the variables, indicating that as one variable increases the other tends to decrease.
4. If the computed value is positive “one” correlation is perfect positive.
5. The values for the Karl Pearson correlation coefficient are from -1 to +1.
6. It is perfect negative correlation.
7. Scatter diagram, Karl Person’s coefficient of correlation
8. High degree of correlation

## Self-Assessment Questions

1. What are the properties of correlation coefficient?
2. Explain the various methods of studying correlation.
3. Write a short note on the concept of probable error.
4. Briefly explain the different degrees of correlation.
5. Explain how Spearman’s rank correlation is computed when actual ranks are given and when actual ranks are not given.
6. What do you mean by coefficient of correlation?
7. What is Attenuation-Corrected Correlation Coefficient?
8. Explain how Spearman’s rank correlation is computed when there are tied ranks.

## 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 |

**Ans: -0.30**

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 |

**Ans: -0.71**

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 |

**Ans: -0.904**

4. Calculate the coefficient of correlation from the following data by the Spearman's Rank Differences method.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Price of Tea (₹): | 75 | 88 | 95 | 90 | 60 | 80 | 81 | 50 |
| Price of Coffee (₹): | 120 | 134 | 150 | 115 | 110 | 140 | 142 | 100 |

**Ans: 0.69**

5. Calculate the Pearson correlation coefficient between income and weight from the following data. Also comment on the result?

| | | | | | | |
|---|---|---|---|---|---|---|
| Income (₹): | 100 | 200 | 300 | 400 | 500 | 600 |
| Weight (Lbs.): | 120 | 130 | 140 | 150 | 160 | 170 |

**Ans: r =1, P.E = 0, r>P.E, r is highly significant**

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 |

**Ans: r =0.95, P.E = 0.022, r>P.E , r>6 times P.E, r is significant**

## Suggested Readings


1. C.R.Kothari (2013). *Quantitative Techniques*. Vikas Publishing House
2. S.P.Guptha-(2012). *Statistical Methods*. Sultan Chand &Sons,New Delhi
3. Riggleman and Frisbee. *Business Statistics*, McGraw Hill Book Co., New York.
4. Secrist Horace. *An Introduction to Statistical Methods*, The Macmillan Co., New York.
5. Simpson and Kafka. *Basic Statistics*, Oxford and I.B.H. Publishing Co., Calcutta.

# Unit - 3
# Regression

## Learning Outcomes

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

- ✓ get a basic knowledge of regression analysis
- ✓ find the equations of regression lines and use them where appropriate
- ✓ use a regression line to predict values of y for the value of x
- ✓ evaluate the use of regression analysis in decision making

## Prerequisites

In a village in Kerala, Vishnu, a paddy farmer, noticed that his rice yield varied every year based on how much water his field received during the growing season. Curious to understand this relationship better, he began recording data: the number of hours he used his irrigation pump each week and the corresponding rice yield at the end of the season. Over time, Vishnu saw a pattern - when irrigation hours increased, his yield also increased, but only up to a certain point. Using regression, Vishnu could analyze this data to predict how much rice he might harvest if he irrigates his fields for a specific number of hours. Here, the irrigation hours are the independent variable (the factor influencing the result), and the rice yield is the dependent variable (the outcome).

By using regression, Vishnu can make smarter decisions about managing his resources. For instance, if he knows how much water is available and the likely rainfall during the season, he can decide how many hours to run his pump to achieve the best yield without wasting water or electricity. This helps him save costs while maximizing production.

This example shows how regression connects variables to provide actionable insights. understanding regression starts with recognizing how one factor can influence another, like water affecting crop yield. Whether it's farming, business, or any other field, regression is a powerful tool that transforms raw data into meaningful predictions, enabling better planning and decision-making

## Keywords

Linear regression, Line of best fit, Dependent variables, Independent variables

## Discussion

### 5.3.1 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. Regression analysis is a type of statistical method that estimate the relationship between a dependent variable and one or more independent variables. In this, independent variable is also known as regressor or predictor or explanatory variable while the dependent variable is also known as regressed or explained variable.

To understand the concept of independent and dependent variables, one should understand the meaning of variables. Variables are defined as" the properties or kinds of characteristics of certain events or objects". A variable is any characteristics, number, or quantity that can be measured or counted. A variable may also be called a data item. Examples of variables are, age, sex, annual income or expenses of a business or household, capital of a company, capital expenditure are examples of variables.

Regression analysis is used in all aspects of life and the need for predictions are the norms for life these days. Let us take the example of business predictions. Imagine a company wants to predict their sales for next months. So, this will be the dependent variable, because the sales value is what we are going to predict. The dependent variables refer to that type of variable that measures the effect of the independent variable(s) on the test units. We can also say that, the dependent variables are the types of variables that are completely dependent on the independent variable(s).

The most frequently used regression analysis is **linear regression**, which involves creating a line of **"best fit"** which can be used for predictions.

## 5.3.2 Types of Regression

We will study four types of regression, linear, non-linear, simple and multiple regressions. Let us explain one by one.

**Linear Regression**

Simple linear regression is a regression model, that estimates the relationship between a single independent variable and a dependent variable using a straight line. Both variables need to be quantitative in nature. Regression is simply establishing a relationship between the independent variables and the dependent variable. But linear regression is establishing a relationship between the features of dependent and independent variables, that can be best represented by a straight line. Linear regression is commonly employed to analyse and understand relationships such as determining the speed of a vehicle, ship, or aircraft based on relevant features. For instance, if we want to investigate the correlation between soil erosion and the amount of rainfall, we can utilize linear regression to examine this relationship.

**Non-linear Regression:** Linear regression is a statistical model that describes the relationship between one or more independent variables and a dependent variable using a linear equation. Nonlinear regression is a statistical model that describes the relationship between one or more independent variables and a dependent variable using a nonlinear equation.

Nonlinear regression is often used to model relationships that are not linear in nature, such as exponential growth, logistic growth, and saturation curves. It is commonly used in a variety of fields, including finance, economics, biology, and physics.
Here are some examples of when nonlinear regression might be used:

- Modeling the growth of a population over time
- Predicting the demand for a product
- Forecasting the performance of a stock

.

**Simple Regression**

Simple regression is a statistical model that describes the relationship between two variables. It can be either linear or nonlinear. "Simple regression" and "linear regression" are terms that are sometimes used interchangeably, but they refer to slightly different concepts in statistics. Linear regression is a specific type of simple regression where the relationship between the two variables is modelled using a straight-line equation. "Simple regression" is a broader term that involves any modelling of the relationship between variables, whether it’s linear or not. E.g., Rent payable vs. square feet of the building.

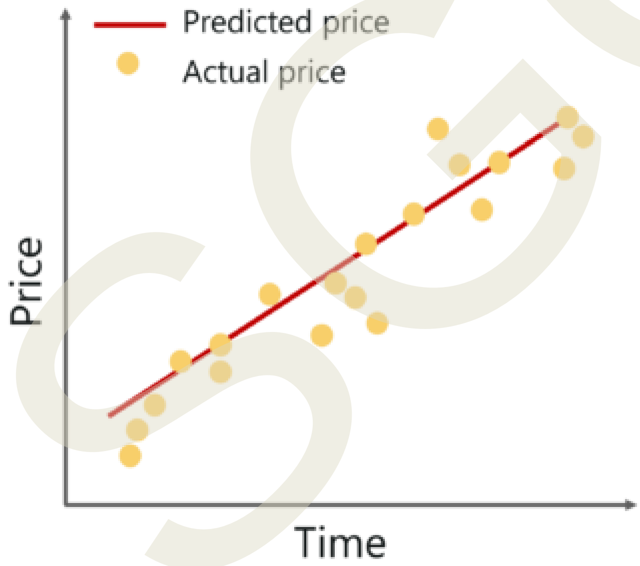
**Multiple Regression**

In multiple regression analysis, we examine scenarios where multiple independent variables collectively influence a single dependent variable. For instance, when investigating the factors that affect rent payments, such as the square footage of the building, available amenities, and maintenance charges, we're dealing with a prime example of multiple regression.

## 5.3.3 Line of best fit

The line of best fit is drawn to represent the relationship between two or more variables. Specifically, it is drawn across a scatter plot of data points to illustrate the relationship between those data points. Regression analysis employs mathematical methods, such as the least squares method, to establish a definitive relationship between predictor variable(s) and the target variable. The least squares method is one of the most effective techniques for drawing the line of best fit.

If we were to plot the best fit line to depict the sales of four-wheelers in India over a period of time, it would resemble something like this:




Notice that the line is as close as possible to all the scattered data points. This is what an ideal best fit line looks like.

## 5.3.4 Distinction Between Correlation and Regression

| Points of Distinction | Correlation | Regression |
|---|---|---|
| Meaning | "correlation" establishes a relationship or connection between the variables. | 'Regression' explains how an independent variable is numerically associated with the dependent variable. |
| Inter dependence between variables | Explains the inter-dependence between the variables | Does not explain the interdependence between the variables |
| Objective | The primary objective of Correlation is to find out a numerical value expressing the association between the values. | Regression's main purpose is to calculate the values of a random variable based on the values of a fixed variable. |
| Computation objective | To establish in one value the degree and direction of relationship between the variables | To compute the value of one variable given the value of another variable which are established as related variables |
| Purpose of computation | To assess the strength and direction of the relationship between two variables. | To predict or explain the value of one variable based on the value of another variable |
| Output | A correlation coefficient (r) with a value between -1 and 1. A positive correlation indicates a positive relationship between the two variables, while a negative correlation indicates a negative relationship. | A regression equation that can be used to predict the value of the dependent variable based on the value of the independent variable. |
| Mutually Dependent Variables | Yes | No. One variable is dependent and another independent |
| Assumptions | No assumptions about causation or the presence of dependent/independent variables | Assumes a causal relationship and defines a clear dependent variable. |
| Direction of Analysis | Bivariate (between two variables) or multivariate (among more than | Typically, univariate (one dependent variable and one or more independent variables). |

| | | |
|---|---|---|
| | two variables). | |
| Example | Measures the relationship between hours of study and exam scores to see if they're correlated. | Predicts exam scores based on hours of study, prior test scores, and attendance |

## Recap

- Regression - mathematical measure of the average relationship between two or more variables in terms of the original units of the data.
- Linear regression - estimates the relationship between one independent variable and one dependent variable using a straight line.
- Non-linear Regression - study more than one independent variable and one dependent variable.
- Simple Regression is the study of relationship between two variables whether it is linear or non-linear.
- Multiple regression - more than one independent variable and one dependent variable.
- Line of best fit - drawn to represent the relationship between two or more variables.

## Objective Questions

1. What statistical technique is commonly used to model the relationship between a dependent variable and one or more independent variables in order to make predictions or understand the impact of the independent variables on the dependent variable?
2. How many variables are involved in a simple linear regression?
3. Which method is used in regression analysis to estimate the model parameter?
4. In regression analysis, what is the primary purpose of the Method of Least Squares?
5. What is the primary difference between simple linear regression and multiple linear regression?
6. What is dependent variable?
7. What is an independent variable?

## Answers

1. Regression
2. One independent variable and one dependent variable
3. Method of least square
4. Minimize the least square error
5. Simple linear regression involves one independent variable, whereas multiple linear regression involves two or more independent variables.
6. The variable being predicted or explained.
7. The predictor variable(s) used for predictions

## Self-Assessment Questions

1. Explain the term Regression.
2. Explain is the difference between correlation and regression analysis?
3. State the importance of line of best fit.
4. Explain the various types of Regression.

## Assignments

1. Explain the concept of simple linear regression. What are the key components of a simple linear regression model, and how is it used in practice?
2. Define the terms "dependent variable" and "independent variable" in the context of simple linear regression. Provide examples of real-world scenarios where simple linear regression can be applied.
3. Describe the difference between correlation and regression. Provide examples of situations where each would be more appropriate to use.
4. Discuss the difference between simple linear regression and multiple linear regression. When would you choose one over the other in an analysis?

## Suggested Readings


1. Dr. P.R. Vittal (2012). *Business Maths & Statistics*. Margham Publications
2. C.R.Kothari (2013). *Quantitative Techniques*. Vikas Publishing House.
3. S.P.Guptha-(2012). *Statistical Methods*. Sultan Chand &Sons,New Delhi
4. C. B Gupta & Vijay Gupta (2004). *An Introduction to Statistical Methods*. Vikas Publishing
5. Goel & Goel (2014). *Mathematics & Statistics*. Taxmann Allied Services Pvt Ltd
6. Sancheti and V.K.Kapoor (2014). *Business Mathematics*. Sultan Chand & Sons
7. Wikes, F.M (1998). *Mathematics for Business*, Finance and Economics. Thomson Learning

# Unit - 4

# Measure of Regression

## Learning Outcomes

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

- ✓ have a basic knowledge of different measures of regression analysis
- ✓ understand and develop statistical applications and analytical skills using regression analysis
- ✓ acquire knowledge on the functioning of correlation and regression analysis
- ✓ gain the applicability of regression in decision making.

## Prerequisites

After acquiring a grasp of regression analysis, we delve into the realm of measure of regression, which serve as critical tools for assessing the accuracy, reliability and effectiveness of regression models. These measures provide us with valuable insights into how well our models fit the data, the strength of relationship between variables, and the precision of our predictions. By understanding and applying measure of regression, we gain the ability to evaluate the practical significance of our models and make informed decisions based on their outcomes.

Regression analysis is a type of statistical method that estimate the relationship between a **dependent variable** and one or more **independent variables**. The most frequently used regression analysis is **linear regression**, which involves creating a line of **"best fit"** which can be used for predictions.

## Keywords

Freehand method, Method of least squares, Regression lines

## Discussion

### 5.4.1 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's 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.

### 5.4.2 Lines of Regression

If two variables in a bivariate distribution are related, you will notice that the points in the scatter diagram tend to cluster around a specific curve, referred to as the 'curve of regression.' In cases where the relationship between the variables is linear, this curve is 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 as '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.

### 5.4.3 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 $x$ forgiven 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 of y on $x$ is expressed as follows;

$$y = a + bx$$

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

$$\sum x = Na + b\sum y$$
$$\sum xy = a\sum y + b\sum y^2$$

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 x based on y.

## 5.4.4 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 means $(\bar{x}, \bar{y})$ of the 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 5.4.1**

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 | $xy$ | $x^2$ | $y^2$ |
|---|---|---|---|---|
| 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 |
| **∑x=40** | **∑y=30** | **∑xy =260** | **$\sum x^2$ =336** | **$\sum y^2$=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;

$30 = 5a + 40b$ ---------------(1)
$260 = 40a + 336b$------------(2)

Multiplying equation 1 by 8 we get;

$240 = 40a + 320b$ -----------(3)
$260 = 40a + 336b$------------(4)

Subtracting equation (4) from (3)

240 = 40a+320b-

260 = 40a+336b

- 20 = 0 -16b

-16b = -20

$$b = \frac{-20}{-16} = 1.25$$

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;

$40 = 5a + 30b$----------(5)

$260 = 30a + 220b$------(6)

Multiplying equation (5) by 6, we get

$240 = 30a + 180b$------(7)

$260 = 30a + 220b$-----(8)

Subtracting equation (8) from (7)

240=30a+180b -

260= 30a+220b

-20 = - 40b

$$b = \frac{20}{40} = 0.5$$

Substituting the value of "b" in equation (5), we get;

$$40 = 5a + 30b$$

40 = 5a +30 x 0.5

40 =5a +15

5a = 40-15

5a =25

a = $\frac{25}{5}$

= 5

the regression line of $x$ on $y$ is $x = 5 + 0.5y$

i.e., regression line of $y$ on $x$ = $y = 1.25x - 4$

regression line of $x$ on $y$ = $x = 5 + 0.5y$

**Illustration 5.4.2**

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** | $xy$ | $x^2$ | $y^2$ |
|---|---|---|---|---|
| 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 |
| $\sum$x = 30 | $\sum$y = 40 | $\sum$xy = 220 | $\sum x^2$ =340 | $\sum y^2$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;

$40 = 5a + 30b$ ----------(1)

$214 = 30a + 220b$-------(2)

Multiplying equation (1) by 6, we get

$240 = 30a + 180b$ ----------(3)

$214 = 30a + 220b$-------(4)

Subtracting equation (3) from (4)

214 =30a+220b –

<u>240 = 30a+180b</u>

-26 = 40b

b = $\frac{-26}{40}$

= -0.65

This value of "b" can be substituted in equation (1), we get the value of "a". That is;

40 = 5a+30b

40 = 5a +30x-0.65

40 = 5a -19.5

5a = 40+19.5

5a = 59.5

a = $\frac{59.5}{5}$

= 11.9

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;

$30 = 50a + 40b$ --------(5)
$214 = 40a + 340b$-------(6)

Multiplying equation (5) by 4 and (6) by 5 we get;

120 = 200a + 160b -------(7)
1070 = 200a + 1700b ------(8)

Subtracting equation (7) from (8)

1070 = 200a + 1700b –

120 = 200a + 160b

950 = 1540b

b = $\frac{950}{1540}$

= 0.617

This value of "b" can be substituted in equation (1), we get the value of "a". That is;

30 = 50a+40b

30 = 50a +40 x 0.617

30 = 50a + 24.68

50a = 30-24.68

50a = 5.32

a = $\frac{5.32}{50}$

= 0.106

Regression equation x on y is given by x =0.106 + 0.617 y

## 5.4.5 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 $\bar{y} = \frac{\sum y}{n}$ , $\bar{x} = \frac{\sum x}{n}$ , $b_{yx} = r\frac{\sigma_y}{\sigma_x}$

Similarly, regression equation $x$ on $y$ can be obtained as follows;
$(x - \bar{x}\ ) = \ b_{xy}\ (y - \bar{y})$

Where $\bar{y} = \frac{\sum y}{n}$, $\bar{x} = \frac{\sum x}{n}$, $b_{xy} = r\frac{\sigma_x}{\sigma_y}$

**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, $b_{xy}$. 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

**σ**x = Standard deviation of $x$ series
**σ**y = Standard deviation of $y$ series

**Regression Coefficient of $y$ on $x$**

The regression coefficient of $y$ on $x$ is represented by$x$ . 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

$\sigma y$ = Standard deviation of $y$ series
$\sigma x$ = Standard deviation of $x$ series

The regression coefficients $b_{yx}$ and $b_{xy}$ can be easily obtained by using the formula

$$b_{yx} = \frac{\sum(x-\bar{x})\ (y-\bar{y})}{\sum(x-\bar{x})^2}$$

$$b_{xy} = \frac{\sum(x-\bar{x})\,(y-\bar{y})}{\sum(y-\bar{y})^2}$$

**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;

$$r = \sqrt{bxy \;\; \text{x} \;\; byx}$$

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.

## 5.4.6 Properties of regression coefficients

- Correlation coefficient is the geometric mean between the regression coefficients
- If one of the regression coefficients is greater than unity (one), 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 change of origin but not scale.
- Regression coefficients are independent of the origin. This means that if a constant is added to or subtracted from all values of the independent or dependent variables, the values of the regression coefficients will not change.
- 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 5.4.3**

From the_following_data, obtain the regression equation of $x$ on $y$ and $y$ on $x$

| | | | | | |
|---|---|---|---|---|---|
| **x** | 10 | 6 | 10 | 6 | 8 |
| **y** | 6 | 2 | 10 | 4 | 8 |

**Solution**

| **x** | **y** | **x – 8** | $(\mathbf{x-8})^2$ | **y – 6** | $(\mathbf{y-6})^2$ | **(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** |

$$n = 5,\ \bar{x} = \frac{\sum x}{n} = \frac{40}{5} = 8\ ,\ \bar{y} = \frac{\sum y}{n} = \frac{30}{5} = 6$$

Regression equation y on x is given by the equation

$$(y - \bar{y}\ ) = b_{yx}\ (x - \bar{x})$$

$$b_{yx} = \frac{\sum (x-\bar{x})\ (y-\bar{y})}{\sum (x-\bar{x})^2}$$

$$= \frac{20}{16} = 1.25$$

$$(y - 6\ ) = 1.25\ (x - 8)$$

y - 6 = 1.25$x$ - 10

$$y = 1.25x - 4$$

Similarly, regression equation $x$ on $y$ is given by the formula

$$(x - \bar{x}\ ) = b_{xy}\ (y - \bar{y})$$

$$b_{xy} = \frac{\sum (x-\bar{x})\ (y-\bar{y})}{\sum (y-\bar{y})^2}$$

$$= \frac{20}{40} = 0.5$$

$$(x - 8\ ) = \ 0.5\ (y - 6)$$

$$x - 8 \ = 0.5y - 3$$

$$x = 0.5\ y + 5$$

**Illustration 5.4.4**

From the_following_data, obtain the regression equation of $x$ and $y$ and yon $x$

| | | | | | | |
|---|---|---|---|---|---|---|
| **X:** | 20 | 22 | 25 | 26 | 27 | 23 |
| **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** |

$n = 6,\ \bar{x} = \frac{\sum x}{n} = \frac{153}{6} = 25.5,\ \bar{y} = \frac{\sum y}{n} = \frac{198}{6} = 33$

Regression equation y on x is given by the equation

$$(y - \bar{y}\ ) = \ b_{yx}\ (x - \bar{x})$$

$$b_{yx} = \frac{\sum(x - \bar{x})\ (y - \bar{y})}{\sum(x - \bar{x})^2}$$

$$= \frac{38}{101.5} = 0.37$$

$$(y - 33\ ) = \ 0.37\ (x - 25.5)$$

$$y = 0.37x - 23.57$$

Similarly, regression equation $y$ on $x$ is given by the formula

$$(x - \bar{x}\ ) = \ b_{xy}\ (y - \bar{y})$$

$$b_{xy}\ = \frac{\sum(x - \bar{x})\ (y - \bar{y})}{\sum(y - \bar{y})^2}$$

$$= \frac{38}{42}\ \ = 0.9$$

$$(x - 25.5\ ) = 0.9\ (y - 33)$$

$$x\ = 0.9\ y - 4.2$$

**Illustration 4.2.21**

Find the following from the data

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Marks in Economics (X):** | 25 | 28 | 35 | 32 | 31 | 36 | 29 | 38 | 34 | 32 |
| **Marks in Statistics(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) | $(x\text{-}32)^2$ | $(y\text{-}38)^2$ |
|---|---|---|---|---|---|---|
| 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** |

1) $n = 10$, $\bar{x} = \frac{\sum x}{n} = \frac{320}{10} = 32$ , $\bar{y} = \frac{\sum y}{n} = \frac{380}{10} = 38$

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

$$(x - \bar{x}) = b_{xy}(y - \bar{y})$$

$$b_{xy} = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(y - \bar{y})^2}$$

$$= \frac{-93}{398} = -0.2337$$

$$(x - 32) = -0.23(y - 38)$$

$$x = -0.23y + 40.74$$

2) Coff. of correlation $r = \sqrt{bxy \times byx}$

$= \sqrt{-0.66 \times -0.23}$

$= \sqrt{0.1518}$

$= \pm 0.389$

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 4.2.21**

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 $4x - 5y + 33 = 0$ -----(1)

$20x - 9y = 107$ -----(2)

$4x - 5y = -33$ ----- (3)

$20x - 9y = 107$ ----(4)

Multiply equation 3 with 5 we get:

$20x - 25y = -165$ -

$\underline{20x - 9y = 107}$

$16y = 272$

$y = \frac{272}{16}$

$= 17$

When y =17, $x$ =

$20x - 9\times17 = 107$

$\underline{20x - 153 = 107}$

$20x = 260$

$x = \frac{260}{20}$

$=13$

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

$$y = \frac{4}{5}x + \frac{33}{5} \text{ we get } b_{yx} = r\frac{\sigma y}{\sigma x} = \frac{4}{5}$$

Similarly, Rewriting the regression lines of x on y, $20x - 9y = 107$ as

$$x = \frac{9}{20}y + \frac{107}{20} \text{ we get } b_{xy} = r\frac{\sigma x}{\sigma y} = \frac{9}{20}$$

Thus

$$r = \sqrt{b_{yx}.b_{xy}} = \sqrt{\frac{4}{5}.\frac{9}{20}} = \pm 0.6$$

Since $b_{yx}$ and $b_{xy}$ are positive, $r = 0.6$

**Illustration 5.4.7**

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 $x = 24$, $y = 10.556 \times 24 + 794.74 = 1048.084$

## Recap

- Method of Least Square-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.
- Regression Coefficients - Regression coefficient of $x$ on $y$ and regression coefficient of $y$ on $x$.
- Correlation coefficient- Geometric mean between the regression coefficients
- Freehand method - Graph is plotted based on the variables, by taking time on X axis and the other variable on Y axis.

## Objective Questions

1. Which is a mathematical method used to find the best fit line that represents the relationship between an independent and dependent variable?
2. If the points in the scatter diagram cluster round a straight line, then what is the name of such curve?
3. Which is the point of intersection of regression lines?
4. If correlation coefficient is zero, then what is the nature of regression lines?
5. What is the primary purpose of a regression line in linear regression analysis?
6. In a simple linear regression model with the equation y = 2x + 3, what does the coefficient 2 represent?
7. What does the term "residual" refer to in regression analysis?

## Answers

1. Method of Least Square
2. Linear Regression lines
3. $(\bar{x}, \overline{y})$
4. regression lines are perpendicular.
5. To make predictions
6. The slope of the regression line
7. The difference between the observed and predicted values

## Self-Assessment Questions

1. Explain the term method of least square.
2. Explain the concept of Regression with real life examples. Why should there be in general two lines of regression for each bivariate frequency distribution?
3. Distinguish between correlation and regression as concepts used in statistical analysis.
4. What are the properties of regression coefficient?

## Assignments

1. The following data relate to the age of husbands (X) and wives (Y). 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 |

**Ans: Regression equation $y$ on x is $y = 0.76\,x + 1.56$**
**Regression equation $x$ on $y$ is $x = 0.82\,y + 12.22$**
**Age of husband, when the age of wife is 25 years = 32.72 i.e., 33 years.**

2. 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?

**Ans: Regression equation $x$ on y is** $x = 2.198y + 67.56$
**Regression equation $y$ on $x$ is** $y = 0.39\,x + 34.65$
**When $x$=70, y = 62.03 crores.**

3. 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 (₹ Thousands) | ₹ 55 | ₹28.7 |
| Profit Earned (₹ Thousands) | ₹. 13 | ₹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 ₹ 50000?
c) Estimate the amount of capital to be employed, if profit earned is ₹ 20000?

**Ans: Regression equation $x$ on y is** $x = 3.24y + 12.86$
**Regression equation $y$ on $x$ is** $y = 0.28\,x - 2.64$
**When $x$=50.000, y = ₹11.57**
**When $y$=20.000, y = ₹77.69**

4. 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?

**Ans: Regression equation $x$ on y is** $x = 1.36y - 5.19$
**Regression equation $y$ on $x$ is** $y = 0.61\,x + 14.81$
**Correlation coefficient r = 0.913**

5. 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.

   **Ans:** $\bar{x} = 3, \bar{y} = \frac{1}{2}, r = -0.33$

6. 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 |

   **Ans:** , $y = 137.3$

## Suggested Readings


1. Chou, Ya-Lun. *Statistical Analysis*, Holt, Rinehart and Winston, New York.
2. Croxton and Cowden. *Applied General Statistics*, Prentice Hall, London and Prentice Hall of India.
3. Croxton and Cowden. *Practical Business Statistics*, Prentice Hall, London.
4. C.R.Kothari (2013). *Quantitative Techniques*. Vikas Publishing House.
5. S.P.Guptha-(2012). *Statistical Methods*. Sultan Chand &Sons,New Delhi
6. C. B Gupta & Vijay Gupta (2004). *An Introduction to Statistical Methods*. Vikas Publishing
7. Goel & Goel (2014). *Mathematics & Statistics*. Taxmann Allied Services Pvt Ltd

# BLOCK 6

# Index Numbers

# Unit - 1
# Index Number

## Learning Outcomes

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

- ✓ develop an understanding on the meaning and characteristics of index number
- ✓ know the limitations of index number.
- ✓ identify the problems in constructing index number

## Prerequisites

In a modern family living in a busy city, the Sharma family started to notice that their expenses were creeping up every month. Mrs. Sharma felt that the grocery bills were higher than usual, while Mr. Sharma thought their utility costs had increased. Their teenage children, Priya and Aarav, also thought that their entertainment expenses had risen. To figure out the truth, the family decided to track their expenses more carefully, but before they could start, they needed to lay down some basic rules.

"First, we need to decide what we are going to track," Mr. Sharma said. "We can't track everything." Mrs. Sharma suggested they focus on the essentials: groceries (like rice, milk, vegetables, and bread), utility bills (electricity, water, and internet), and transportation costs (fuel and public transport). Priya added, "Let's also include our entertainment costs, like streaming subscriptions and dining out." Once they agreed on the categories, they needed reliable data. Aarav said, "Let's check our past grocery receipts, utility bills, and transport expenses for the last six months." They also agreed to choose a base year to compare everything. "Let's use last year as our base year," suggested Mr. Sharma, "since it was a normal year without any big changes."

Finally, they decided to use simple percentages to calculate how much their costs had increased. "Percentages are easy to understand," Mrs. Sharma said, "and it'll show us exactly how much more we're spending." With clear goals, accurate data, and a simple calculation method, the Sharma family created a system to track their expenses. This helped them understand where their money was going and make smarter financial decisions for the future.

## Keywords

Index number, Price index, Quantity index

## Discussion

### 6.1.1 Introduction

Index numbers offer a standardized and concise way to monitor changes and trends in economic, financial, or social indicators over time. By establishing a base period or reference value, they facilitate comparisons and provide valuable insights into how these variables evolve. This simplifies complex data and allows for smarter decision-making and analysis.

An index number is a device used to measure the relative change in the magnitude of a group of related variables in different situations. The situations may be two different periods of time or places and the group of related variables may be prices or quantities. It represents the general trend of diverging ratios, from which it is calculated. It is a measure of the average change in a group of related variables over two different situations. The comparison may be between categories such as persons, schools, hospitals etc. An index number also measures changes in the value of the variables such as prices of specified list of commodities, volume of production etc. In general, index numbers are used for comparison of changes in the prices or quantities of different periods of time. When comparison is made between prices of two periods, it is called index number of prices and when the comparison is made between quantities, it is called index number of quantities. The year for which the comparison is made is the base year.

### 6.1.2 Definition

**John I. Griffin** has described it as "An index number is a quantity which by reference to a base period, shows by its variation, the changes in the magnitude over a period of time." Index numbers are the numbers which express the value of a variable at any time (current period) as a percentage of the value of that variable at some reference period or base period.

**Edge worth** gave the definition of index numbers as- "Index Number shows by its variation the changes in a magnitude which is not susceptible of either accurate measurement in itself or of direct variation in practice".

**Meaning**

Index numbers are statistical tools that provide a simplified representation of complex data, making it easier to analyse trends, compare values, and draw meaningful conclusions. These numerical indicators are widely used in economics, finance, and various fields to measure changes in variables over time or across different categories.

## 6.1.3 Characteristics of index numbers

- **Relative Measurement**

Index numbers measure changes in relation to a base period. They make a comparison between the variable values in subsequent periods and the base period. This relative measurement makes it possible to analyse changes over time and identify trends.

- **Simplification**

Index numbers reduce complex data to a single measure, hence simplifying it. Index numbers give a summary of the general change in the variable rather than providing various data points. This simplification makes it easier to comprehend and evaluate the data.

- **Comparison and Benchmarking**:

Index numbers allow data to be compared across time periods, geographies, or industries. They serve as a standard or reference point against which changes in variables can be measured. This comparison assists in understanding the relative performance or changes of several entities or time periods.

- **Tracking Trends**

Index numbers are helpful instruments for tracking variable patterns. It is possible to identify patterns, directional changes, and magnitudes of the variable being monitored by analysing changes in index numbers across time. This tracking is useful for forecasting and making decisions.

- **Quantitative Representation**

Index numbers are stated as percentages or ratios, and they provide a quantitative depiction of variable changes. This numerical representation makes the magnitude of changes easier to understand and analyse. It enables for more precise analysis and makes comparisons easier.

- **Standardised Measurement**:

Index numbers give a standardised measurement for evaluating variable changes. They provide a metric that may be applied uniformly across multiple sources or contexts. This standardisation provides data analysis uniformity and comparability.

- **Reflecting Weighted Importance**:

Index numbers can incorporate weighting schemes to reflect the importance of different variables within the index. Weighting assigns greater significance to certain variables, ensuring that they have a larger impact on the overall index. This consideration of weights ensures a more accurate representation of the variable being measured.

- **Application in Various Fields**:

Index numbers are used in a variety of sectors, including economics, finance, labour markets, demography, and more. They are used to measure inflation rates, analyse stock market performance, assess purchasing power, and make informed decisions based on accurate data. Index numbers are useful in a variety of fields due to their adaptability.

## 6.1.4 Importance of index numbers

Earlier index numbers were used to track changes in commodity prices. However, the use of index numbers has expanded significantly in recent years. They are presently used to measure changes in physical quantities, volume of production, volume of trade, cost of products and services, and share and stock prices. Index numbers play a crucial role in various fields, including economics, finance, statistics, and social sciences. These statistical measures serve as essential tools for analysing and interpreting data, tracking changes over time, and making informed decisions.

- **Economic Indicators**

Index numbers are critical in measuring economic indicators such as inflation, economic growth, and employment rates. The Consumer Price Index (CPI) measures changes in the average price of goods and services, assisting policymakers in monitoring inflation rates and adjusting monetary policies. GDP and associated indexes provide information on a country's economic performance and growth.

- **Comparison across Time and Space**

Index numbers allow for meaningful comparisons of data across different periods, regions, or variables. They provide a standardized framework to measure changes over time or compare different entities. This comparative aspect enables researchers, policymakers, and businesses to assess trends, identify variations, and make informed decisions.

- **Development Assessment**

Index numbers are critical for analysing and comparing levels of development across countries or regions. The Human progress Index (HDI) measures a country's overall progress by combining factors such as life expectancy, education, and income.

Policymakers can use the HDI to identify areas for improvement, prioritise development activities, and track progress over time.

- **Social Science Research**

Index numbers are used in social science research, notably surveys and studies aimed at quantifying subjective experiences. Researchers can analyse and compare large-scale survey data by employing indices to quantify characteristics such as satisfaction, happiness, or well-being. Index numbers make it easier to see patterns, correlations, and social trends, which improves understanding of human behaviour and societal dynamics.

- **Business and Marketing Insights**

In business and marketing, index numbers are used to assess consumer happiness, brand loyalty, and market competitiveness. These indices provide vital insights to businesses, assisting them in making data-driven decisions to improve goods, optimise marketing tactics, and improve customer experiences.

## 6.1.5 Limitations

Index numbers are versatile statistical tools widely used in economics, finance, and various fields to simplify complex data and make meaningful comparisons. However, they have a number of drawbacks that may compromise their relevance, accuracy, and interpretation

- **Base Period Dependency**

Index numbers are typically calculated relative to a base period. The choice of this base period can impact the results significantly. The base period selection has a considerable bearing on the outcomes. It is difficult to compare indices calculated using various basis periods since they can provide different index values. Because of this restriction, the analysis may be biased in terms of time.

- **Composition Bias**

Index numbers are frequently used to indicate a defined set of things or components. If the content of this basket varies over time, the index may become biased. A Consumer Price Index (CPI), for example, may not effectively reflect the cost of living if the items in the basket do not correspond to consumer spending patterns.

- **Substitution Bias**

Traditional index numbers may not properly account for customer substitution behaviour. When the price of one item in the basket rises dramatically, consumers may switch to cheaper alternatives. Traditional indexes may not effectively capture this real-world customer behaviour.

- **Quality Changes**

Index numbers often do not account for changes in the quality of goods and services. For example, improvements in the quality of smartphones or automobiles may not be adequately reflected in an index.

- **Weighting Issue**

The weights assigned to different components in an index can be somewhat subjective. These weights may not always reflect the true importance or relevance of these components to the population being measured. Inaccurate weighting can skew the index's representation of reality.

- **Inflationary Bias**

Price indices, like the CPI, may overestimate or understate inflation for a variety of reasons, including the challenge of precisely tracking price changes and accounting for changing product quality. Decisions about economic policy may be significantly impacted by this.

- **Lack of Regional Variation**

Some indices may fail to account for regional variations in prices and consumption habits. This constraint is especially relevant in countries or areas with varying economic conditions and lifestyles.

- **Data Availability**

The precision of index numbers is strongly reliant on the availability of trustworthy data. Data restrictions can have an impact on the index's precision, especially when full data is unavailable.

- **Assumption of Constant Weights**

Many indices are assumed to have constant weights across time. In reality, the relevance of various components may vary and consumer preferences can shift. Constant weights may not effectively represent these dynamics.

- **Complexity**

Index numbers can be difficult to calculate and interpret, limiting their accessibility and utility, particularly for the general public.

## 6.1.6 Problems in construction of index numbers

The generation of index numbers is a fundamental method for summarising and comparing data in economics, statistics, and other domains. It is not, however, without hurdles and possible problems. Changes in values over time, across categories, or between distinct data points are represented by index numbers. They are useful for identifying trends, comparing data, and making decisions. Regardless of its utility, index number construction can be a complex process, with various issues that must be carefully addressed to assure the accuracy, relevance, and dependability of the final indices. In this section, we look deeper into the issues surrounding index number production.

- **Selection of Base Year**

One of the most important decisions in index construction is selecting a base year or period against which other observations will be evaluated. The base year acts as a reference point, and choosing an incorrect base year might have a substantial impact on the outcomes. Using an out-of-date base year may not adequately reflect current economic conditions, while using a recent base year may miss long-term trends. The base year should preferably be current while also being reflective of average conditions.

- **Weighting Issues**

It is critical to assign weights to distinct index components because these weights indicate the relative relevance of each component in the overall index. Choosing adequate weights, on the other hand, can be difficult. Weights should ideally be based on current and precise data; however, this information is not always readily available. Furthermore, customer tastes and economic conditions can shift over time, making constant weights less useful.

- **Data Quality**

The dependability and correctness of the data used to generate an index are critical. Data quality difficulties can originate from a variety of sources, such as sample errors, inaccuracies in reporting, or limits in data collection methods. Data that is inconsistent or biased might result in inaccurate index values and misrepresentations of economic reality.

- **Composition Bias**

Many index numbers indicate a predetermined set of things or components referred to as the "basket." This basket's composition is typically determined by historical consumption patterns. If the composition does not closely reflect actual consumption trends, the index may be biased. A Consumer Price Index (CPI), for example, may not effectively reflect the cost of living if the items in the basket do not correspond to what people are currently purchasing.

- **Substitution Bias**

Consumer substitution behaviour may not be fully accounted for by traditional index numbers. In practise, if the price of one item in the basket rises dramatically, consumers may opt for less expensive alternatives. Conventional indexes may not effectively represent this real-world consumer behaviour, causing inflation to be overstated.

- **Quality Changes**

The quality of goods and services can alter throughout time. For example, technological developments may result in higher product quality or durability. Many index numbers, however, do not take these quality gains into consideration. This constraint might lead to an underestimation of actual economic growth or an overestimation of inflation.

- **Basket Updating**

To keep an index relevant, the basket components should be updated on a regular basis to reflect current consumption patterns. The process of picking items for the basket and determining their weights, on the other hand, might be subjective and may not necessarily reflect consumers' genuine preferences. Furthermore, upgrading the basket raises the issue of guaranteeing consistency and comparability with historical data.

- **Inflationary Bias**

Due to a variety of reasons, price indices such as the CPI may overestimate or understate inflation. Among these factors are the difficulties in accounting for changing product quality, accurately quantifying pricing fluctuations, and recording shifts in consumption habits. Price measuring precision is critical for making educated economic policy decisions and determining pay adjustments.

- **Geographic Variation**

In some circumstances, index numbers may fail to account for regional variations in prices and consumption patterns. This constraint is especially relevant in countries or areas with varying economic conditions and lifestyles. Failing to consider these variations can lead to misleading or inadequate assessments of economic conditions.

- **Data Availability**

The availability of extensive and up-to-date data determines an index's accuracy and reliability. In some circumstances, data may be incomplete or missing, making it difficult to create a reliable index. This is especially important in developing countries or places with poor data infrastructure.

- **Temporal Variation**

The importance and usefulness of index components may fluctuate throughout time. For example, the importance of specific commodities and services in the consumer basket may change as a result of technological achievements or changing societal requirements. Failure to account for such temporal fluctuations may result in a less relevant index.

- **Assumption of Constant Weights**

Many indices assume constant weights across time, implying that the relative importance of distinct components remains constant. However, the importance of individual components can fluctuate due to changes in consumer tastes, technical improvements, or economic trends. Constant weights may not effectively reflect these dynamics.

- **Lack of Transparency**

Transparency during the construction process is essential for user confidence and understanding. Failing to document and communicate the methodology, data sources, and assumptions used in creating an index can hinder users' ability to assess its limitations and potential biases.

## 6.1.7 Types of index number

Index numbers are statistical tools that are used to measure changes in economic and noneconomic variables across time. They are commonly used in economics, finance, and other fields to compare the relative changes in distinct quantities, prices, or values. There are various sorts of index numbers, each with a distinct function. Here are some examples of index numbers.

- **Price Index**

A Price Index is a sort of index number that is used to track price changes in a certain basket of products and services across time. It measures the degree of inflation or deflation in an economy by comparing current prices to those in a base period. The Consumer Price Index (CPI), which analyses changes in the cost of living for consumers, is one of the most well-known price indices. Price indices are critical tools for tracking inflation, adjusting salaries, and making sound economic policy decisions.

.

- **Quantity Index**

Quantity Index numbers are concerned with changes in the physical quantities or volumes of products and services, rather than their prices. These indices are often used to analyse production trends in industries such as manufacturing and agriculture. A quantity index, for example, could measure the output of a certain crop or the production of manufactured goods through time. These indexes are useful for analysing changes in physical output, which can be critical for resource allocation and economic planning.

- **Value Index**

A Value Index uses both price and quantity information to calculate changes in the total value of a certain set of goods and services. These indices are critical for evaluating general economic trends. For example, the Gross Domestic Product (GDP) deflator is a value index that measures changes in an economy's total price level. Value indices provide a full view of economic performance by taking into account both price and quantity changes, making them useful tools for policymakers and analysts.

## Recap

- Index number - measure changes in relation to a base period
- Index number - reduce complex data to a single measure
- Basket components should be updated on a regular basis
- Price Index - a sort of index number that is used to track price changes
- Quantity Index number- changes in the physical quantities or volumes of products and services,
- Value Index - both price and quantity information to calculate changes in the total value
- Index numbers -give a standardised measurement for evaluating variable changes

## Objective Questions

1. What is the primary purpose of index numbers in measurement?
2. In what form are index numbers typically expressed?
3. In which field are index numbers used to quantify subjective experiences such as happiness or well-being?
4. Which bias may occur if the composition of the basket in an index does not reflect actual consumption trends?
5. What does the term "composition bias" in index numbers refer to?
6. What bias may occur if an index does not consider regional variations in prices and consumption patterns?
7. What is the primary purpose of a Consumer Price Index (CPI)?

## Answers

1. To simplify complex data
2. Percentage or ratios
3. Social science research
4. Substitution bias
5. Failing to account for consumer substitution behaviour
6. Geographic variation
7. To measure changes in the average price of consumer goods and services over time.

## Self-Assessment Questions

1. What is the primary purpose of index numbers in statistics?
2. How do index numbers simplify complex data in various fields?
3. Why is the selection of a base period important in index number calculations?
4. Describe the role of index numbers in tracking trends.
5. What is the importance of reflecting weighted importance in index numbers?
6. Name some fields where index numbers are widely used and provide examples of their applications.
7. What are the limitations associated with the use of index numbers?
8. Discuss the importance of transparency in the construction of index numbers.
9. Provide examples of fields where index numbers find practical applications.
10. Explain the importance of data availability in ensuring the accuracy of index numbers

## Assignments

1. Describe the challenges associated with assigning weights to index components.
2. Investigate the role of data quality in the construction of index numbers and provide examples of data quality issues and their impact on index values.
3. Conduct a case study on composition bias in a specific index and identify instances where the composition of the "basket" did not accurately reflect consumer trends and the resulting implications.
4. Research and explain the concept of substitution bias in index numbers and analyse real-world scenarios where substitution behaviour was not fully accounted for.
5. Explore how quality changes in goods and services can affect index numbers and provide examples of industries where quality improvements are significant and discuss potential consequences for index calculations.
6. Evaluate the challenges and subjectivity involved in updating the components of a consumption basket in an index and discuss the importance of maintaining consistency and comparability with historical data.

\

## Suggested Readings


1. C.B Gupta and Vijay Gupta (2004) *An Introduction to Statistical Methods*. Vikas Publishing House.
2. S.P. Gupta. *Statistical Methods*. Sultan Chand and Sons, New Delhi.
3. Frederick E Croxton, Dudley J Cowden and Sidney Klein. *Applied General Statistics*. Prentice Hall India.
4. Naval Bajpai. *Business statistics*. Pearson Educational Publications
5. Dr. S M Shukla and Dr. Sahai. *Principles of statistics*. Sahitya Bhavan Publication, Delhi

# Unit - 2
# Price index number

## Learning Outcomes

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

- ✓ explain the concept of price index
- ✓ get an awareness on the wholesale price index number
- ✓ get an idea on the methods of construction of index numbers

## Prerequisites

In a busy market town, there was a friendly shopkeeper named Ramesh. Ramesh owned a small grocery shop that sold everything from rice and flour to vegetables and spices. Over time, he noticed that his customers were complaining about rising prices. Some customers believed that prices were going up quickly, while others felt they were the same. Ramesh wanted to find out the truth, so he decided to create a simple way to track how much prices were changing. But before he could get started, he realized he needed to follow some basic steps.

"First, I need to decide what to measure," Ramesh thought to himself. "I can't track everything in my shop." After some thought, he decided to focus on a few key items that everyone buys regularly, like rice, wheat flour, sugar, and vegetables. He also realized that he needed reliable data. "I'll gather prices from my shop's records over the past few months to see if there have been any price changes," he said. But then, Ramesh remembered he needed a starting point, or a base year, to compare the prices. "I'll use last year's prices as the base, since it was a normal year without any sudden changes in the market," he decided.

Finally, Ramesh thought about how to calculate the price changes. "I should use percentages because they are easy to understand and will show exactly how much prices have gone up or down." With these simple rules in mind - deciding what to measure, gathering accurate data, choosing a base year, and using percentages - Ramesh set up a way to track the prices in his shop. This helped him understand how the cost of goods had changed and allowed him to make better decisions about pricing in the future.

## Keywords

Price index number, Whole sale price index number, Consumer price index number

## Discussion

The most common and widely used type of price index is the price index number. Price index, a measure of relative price changes composed of a series of numbers structured in such a way that comparing the values for any two periods or places reveals the average change in prices between periods or the average difference in prices between places. Price indexes were initially developed to measure changes in the cost of living in order to calculate the pay increases required to maintain a consistent standard of living. They are still widely used to predict price changes over time and to quantify cost variations across different places or countries. Price index measures the changes in the price level of the current period on the basis of the price level of the base period.

Price index = (Current year price / Base year price) x 100

$$P_{01} = \frac{P_1}{P_0} \times 100$$

Here, $P_{01}$ is the current year's price index based on the base year price of a given commodity, and 100 is the common denominator.

### 6.2.1 Types of Price Index

**Wholesale Price Index Number**

Index numbers are indicators that indicate changes in commodity prices, industrial production, sales, imports and exports, cost of living, and so on during a given time period. These indicators are useful for reviewing and managing current economic positions as well as formulating plans. Some of the most important indices, such as the Wholesale Price Index (WPI), Index of Industrial Production (IIP), Consumer Price Index (CPI), and others, provide a strong indication of what is going on in the economy. WPI is a key indicator for measuring the dynamic movement of wholesale pricing. Prices are constantly shifting in a dynamic world. WPI is used as a deflator for a variety of nominal macroeconomic indicators, including GDP. WPI-based inflation estimates are also used by the government in the formation of trade, fiscal, and other economic policies. WPI is also utilised in the delivery of raw materials, machinery, and construction activities as an escalation clause. Price adjustment (escalation) provisions in long-term sales and purchase contracts are frequently used by businesses seeking effective methods of dealing with price increases. WPI is commonly regarded as a useful objective indexing method in price adjustment clauses by businesspeople, economists, statisticians, and accountants. As a result, the Wholesale Price Index represents the price of a wholesale products basket.

WPI is concerned with the price of items transferred between corporations. It does not focus on things purchased by consumers. WPI's primary goal is to monitor price fluctuations that reflect demand and supply in manufacturing, construction, and industry. WPI assists in measuring an economy's macroeconomic and microeconomic circumstances.

**Retail Price Index**

It measures changes in a sector's or community's price level. Retail prices and quantities of a set of commodities used by the community's residents are collected for its calculation. The retail price index includes the consumer price index or cost of living index.

## 6.2.2 Methods of Construction of Index Numbers

There are two broad groups of index numbers i.e., Simple and Weighted. Both of these methods are further subdivided into Aggregate and Average relatives' methods.

**I. Simple Index Numbers.**

Each item is expected to have the same weight or importance in simple index numbers. There are two types of simple index numbers: Simple Aggregate Method and Simple Average Relatives Method.

**a. Simple Aggregate Method**

This is the simplest way to make index numbers. In this method, the total current year prices of the commodities are divided by the total base year prices, and the quotient is multiplied by 100.

$$P_{01} = \frac{\sum P_1}{\sum P_0} \times 100$$

Here, $P_{01}$ is the Price Index, i.e., the current year's price index on the basis of the price of the base year.

$\sum p_1$ = Total prices of the current year of various commodities.
$\sum p_0$ = Total prices of the base year.

**Steps**

1. To get $\sum p_1$, add the current year prices of various commodities.
2. Calculate $\sum p_0$ by adding the base year prices of several commodities.
3. To acquire the index number, divide $\sum p_1$ by $\sum p_0$ and multiply the quotient by 100.

**Limitations of the method**

The simple Aggregate method is highly defective, despite its ease of calculation. The following are some of the drawbacks:

1. When using this method, the items with the highest unit price significantly impact the index number.
2. The relative importance of goods is not taken into account.

Simple (Unweighted) Aggregate Method

**Illustration 6.2.1**

From the following data calculate Index Number by Simple Aggregate Method.

| Commodity | A | B | C | D |
|---|---|---|---|---|
| Price in 1990 ( ₹) | 160 | 258 | 250 | 139 |
| Price in 1991(₹) | 170 | 165 | 190 | 144 |

**Computation of price index number**

| Commodity | Price (in Rupees) | |
|---|---|---|
| | 1990($p_0$) | 1991($p_1$) |
| A | 160 | 170 |
| B | 258 | 165 |
| C | 250 | 190 |
| D | 139 | 144 |
| Total | $\sum p_0 = 807$ | $\sum p_1 = 669$ |

The price index number using Simple Aggregate Method is given by:

$$P_{01} = \frac{\sum P_1}{\sum P_0} \times 100$$

$= \frac{669}{807} \times 100$

$= 82.90$

**b. Simple Average Relatives Method**

The price relatives of each product are computed first in this method. It is calculated by dividing each commodity's current year price by its base year price and multiplying the quotient by 100.

$$I = \frac{\sum P_1}{\sum P_0} \times 100$$

Where, I is the price relatives of a product.

The arithmetic average of price relatives is used to get the price index for the entire group.

$$P_{01} = \frac{\sum \mathbf{I}}{\mathbf{N}}$$

Following are the advantages and disadvantages of Simple Average Relative Method:

**Advantages:**

a) It is easy to understand.
b) Any average can be used to calculate it.
c) It offers equal weight to all of the items under consideration.
d) It is unaffected by the relative prices of various products.

**Disadvantages:**

a) It is difficult to calculate when the geometric mean is used.
b) It makes it difficult to choose a suitable average.
c) It presumes that all relatives are equally important, which is highly unacceptable from an economic standpoint.

**II. Weighted Index Numbers.**

The index number calculated after assigning appropriate weightage to the various items under consideration is called the weighted index number. Weights are assigned explicitly in this scenario, either based on quantities or the values of the products consumed or on some other rational basis. Creating a usable index number necessitates a deliberate effort to assign each commodity a weight based on its relevance.

### 6.2.3 Types of Weighted Index Numbers

**a. Weighted Aggregate Index Numbers.**

These indices are of the simple Aggregate form, with the exception that weights are set to the various items included in the index. On the basis of different weighting methods, various way of creating index numbers have evolved. The various methods are labelled with the names of the people who proposed them. The methods are as follows:

i. Laspeyre's Method
ii. Paasche's Method
iii. Fisher's Ideal Method
iv. Marshall- Edgeworth Method
v. Kelley's Method

**i. Laspeyre's Method**

In 1871, German economist Etienne Laspeyre developed this approach for constructing price indices. The method assumes that the quantities consumed in the base year and current year are the same and that weights are determined by the quantities consumed in the base year. The formula is:

$$P_{01}(L) = \frac{\sum p1q0}{\sum p0q0} \times 100$$

**Steps**

1. Calculate p1q0 by multiplying the current year price of each commodity by the corresponding base year quantity.
2. Calculate p0q0 by multiplying the base year price of each commodity by its base year quantity.
3. To calculate the price index, divide p1q0 by p0q0 and multiply the quotient by 100.

**Advantages**

The main advantages of the method are:

a) It is simple to understand and compute.
b) It is based on fixed weights since the quantity of the base year is used as the weight of the item in both years.
c) It does not necessitate quantities consumed in the current year.

**Disadvantages**

Following are the main disadvantages of the method:

a) The concept that the same quantities are utilized in the base year and current year is inappropriate.
b) It does not allow the use of averages such as geometric mean, median, etc.
c) This method does not satisfy the ideal index number test.
d) It does not use current year quantities, even though they are accessible.

### ii. Paasche's Method

Paasche, a German mathematician, developed this method in 1874 to improve Laspeyre's method. He used the current year's quantities as the weights of the items in this method. As a result, Paasche provides the following formula:

$$P_{01}(P) = \frac{\sum p1q1}{\sum p0q1} \times 100$$

**Steps**

1. To get $\sum p_1q_1$, multiply the current year prices of various commodities by their respective current year quantities.
2. Calculate $\sum p_0q_1$ by multiplying the base-year prices of various commodities by their respective current year quantities.
3. Multiply the quotient by 100 after dividing $\sum p_1q_1$ by $\sum p_0q_1$.

**Advantages**

a) It is easy to calculate and understand.
b) It is based on constant weights for both the current and base years.
c) The base year quantity is not required.
d) It meets the unit test for the adequacy of an index number formula.

**Disadvantages**

a) The method assumes that consumption quantities are the same in the base year and the current year.
b) Each time an index number is constructed, weights must be determined.
c) The time-reversal, factor reversal, and circular tests required for an ideal index number are not met.
d) It completely ignores the actual quantities consumed in the base year.

### iii. Fisher's Ideal Index Number

The method for constructing index numbers proposed by Prof. Irving Fisher is fisher's ideal index. It is the geometric mean of the index numbers of Laspeyre's and Paasche. The formula is:

$$P_{01}(F) = \sqrt{\frac{\sum p1q0}{\sum p0q0} \times \frac{\sum p1q1}{\sum p0q1}} \times 100$$

**Characteristics**

a) It is based on the Geometric Mean, which is thought to be the ideal average for building index numbers.
b) It satisfies time and factor reversal tests.
c) It removes the bias associated with Laspeyre's and Paasche's methods by considering both the current year's and base year's weights.

**Merits**

a) It is created using the Geometric Mean, the most appropriate average for an index number.
b) It uses all available data, including $p_1$, $p_0$, $q_1$ and $q_0$.
c) It is ideal since it meets time reversal, factor reversal, and unit tests.
d) It reflects the impact of the current year and the base year.

**Demerits**

a) Calculation is quite complex.
b) Each time an index is constructed, it requires data on the current price and quantity.
c) It's a combination of Laspeyre's and Paasche's methods.

**Marshall-Edgeworth Index (MEI)**

The Marshall-Edgeworth Index (MEI) is a price index that seeks to measure changes in the prices of a fixed basket of goods and services over time

$$P01\ (ME) = \frac{\sum p1q0 + \sum p1q1}{\sum p0q0 + \sum p0q1} \times 100$$

**Illustration 6.2.2**

From the following data calculate price index numbers for 1980 with 1970 as base by (i) Laspeyre’s method, (ii) Paasche’s method, (iii) Marshall- Edgeworth method, and (iv) Fisher’s ideal method.

| Commodities | 1970 | | 1980 | |
|---|---|---|---|---|
| | Price | Quantity | Price | Quantity |
| A | 20 | 8 | 40 | 6 |
| B | 50 | 10 | 60 | 5 |
| C | 40 | 15 | 50 | 15 |
| D | 20 | 20 | 20 | 25 |

a. It is stated that Marshall- Edgeworth index number is a good approximation to Fisher's ideal index number. Verify this for the data in Part(a).

**Solution**

Calculations for Price Indices by different formulae

| Commodities | 1970 | | 1980 | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $p_0$ | $q_0$ | $p_1$ | $q_1$ | $p_0q_0$ | $p_0q_1$ | $p_1q_0$ | $p_1q_1$ |
| A | 20 | 8 | 40 | 6 | 160 | 120 | 320 | 240 |
| B | 50 | 10 | 60 | 5 | 500 | 250 | 600 | 300 |
| C | 40 | 15 | 50 | 15 | 600 | 600 | 750 | 750 |
| D | 20 | 20 | 20 | 25 | 400 | 500 | 400 | 500 |
| | | | | | $\sum p_0q_0=$ 1660 | $\sum p_0q_1=$ 1470 | $\sum p_1q_0=$ 2070 | $\sum p_1q_1=$ 1790 |

(i) Laspeyre's Price Index

$$P_{01}^{La} = \frac{\sum p1q0}{\sum p0q0} \times 100 = \frac{2070}{1660} \times 100 = 1.24699 \times 100 = 124.699$$

(ii) Paasche's Price Index

$$P_{01}^{Pa} = \frac{\sum p1q1}{\sum p0q1} \times 100 = \frac{1790}{1470} \times 100 = 1.2177 \times 100 = 121.77$$

(iii) Marshall- Edgeworth Price Index

$$P_{01}^{ME} = \frac{\sum p1q0+\sum p1q1}{\sum p0q0+ \sum p0q1} \times 100 = \frac{2070+1790}{1660+1470} \times 100 = \frac{3860}{3130} \times 100 = 123.32$$

(iv) Fisher's Price Index

$$P_{01}^{F} = \sqrt{\frac{\sum p1q0}{\sum p0q0} \times \frac{\sum p1q1}{\sum p0q1}} \times 100 = \sqrt{\frac{2070}{1660} \times \frac{1790}{1470}} \times 100$$

$$= \sqrt{1.24699 \times 1.2177} \times 100 = \sqrt{1.51846} \times 100$$

$$= 1.23226 \times 100 = 123.23$$

Alter:

$$P_{01}^{F} = \sqrt{P01La \times P01Pa} = \sqrt{124.699 \times 121.77} = \sqrt{15184.597} = 123.23$$

(b) Since $P_{01}^{ME}$ = 123.32 and $P_{01}^{F}$ = 123.23, are approximately equal, Marshall-Edgeworth index number is a good approximation to Fisher's ideal index number.

**Kelly's Price Index**

$$P_{01}^{K} = \frac{\sum q\ p1}{\sum q\ p0} \times 100$$

**Illustration 6.2.3**

Calculate the weighted price index from the following data;

| Materials required | Unit | Quantity required | Price (Rs.) | |
|---|---|---|---|---|
| | | | 1963 | 1973 |
| Cement | 100 lb | 500 lb. | 5.0 | 8.0 |
| Timber | c.ft. | 2,000 c.ft. | 9.5 | 14.2 |
| Steel sheets | cwt. | 50 cwt. | 34.0 | 42.20 |
| Bricks | per'000 | 20,000 | 12.0 | 24.0 |

**Solution.**

Since the quantities (weights) required of different materials are fixed for both the base and current years, we will use Kelly's formula for finding out price index.

Further, for cement unit is 100 lbs. and the quantity required is 500 lbs. Hence, the quantity consumed per unit for cement is 500/100 = 5. Similarly, the quantity consumed per unit for bricks is 20,000/1,000 = 20.

**Illustration 6.2.4**

Computation of Kelly's Index Number

| Materials required | Unit | Quantity required | q | Price (Rs.) | | $q\,p_0$ | $q\,p_1$ |
|---|---|---|---|---|---|---|---|
| | | | | 1963 | 1973 | | |
| | | | | $p_0$ | $p_1$ | | |
| Cement | 100 lb | 500 lb. | 5 | 5.0 | 8.0 | 25 | 40 |
| Timber | c.ft. | 2,000 c.ft. | 2000 | 9.5 | 14.2 | 19,000 | 28,400 |
| Steel sheets | cwt. | 50 cwt. | 50 | 34.0 | 42.20 | 1,700 | 2,100 |
| Bricks | per'000 | 20,000 | 20 | 12.0 | 24.0 | 240 | 480 |
| | | | | | | $\sum qp_0$=20,965 | $\sum qp_1$=31,020 |

Kelly's Price Index is given by: $P_{01}^{K} = \frac{\sum q\,p1}{\sum q\,p0} \times 100 = \frac{31020}{20965} \times 100 = 147.96$

## 6.2.4 Consumer price index number

A price index is a measure of the proportionate, or percentage, changes in a set of prices over time. A consumer price index (CPI) measures changes in the cost of goods and services purchased by households. Such changes have an impact on the real purchasing power of consumers' earnings and their well-being. A price index is normally assigned a value of unity, or 100, in some reference period, and the index values for other periods of time are intended to show the average proportional, or percentage, change in prices from this price reference period. Price indices can also be used to compare price levels among cities, regions, or countries at the same time.

Items contributing to consumer price index number are generally brought under five major groups

a) Food
b) Clothing
c) Fuel and lighting
d) Housing
e) Miscellaneous

**Uses of consumer price index number**

a) The cost-of-living index number is used to adjust dearness allowance in order to maintain the same level of living as in the base date.
b) It is used to determine wage policy, taxation policy, and a variety of other economic issues.
c) It is used to measure money's purchasing power. Money's purchasing power varies inversely with the cost-of-living index.
d) It is employed in income and value deflation. Real income is calculated by dividing real income received during a period by the period's cost of living index.

**The main steps which are required for construction of CPI are described below.**

**(l) Decision about the class of people for whom the index is meant.**
It is essential to determine clearly which group of people the index is intended for, such as industrial workers, instructors, officers, and so on. The index's scope must be clearly stated. For example, when we talk to teachers, we are referring to primary teachers, middle class teachers, etc. or to all the teachers taken together. Along with the class of people it is also necessary to decide the geographical area covered by the index.

**(2) Conducting, family budget enquiry.**
After defining the scope of the index, the following stage is to perform a family budget survey of the population group for which the index is to be built. The goal of a family budget investigation is to establish how much an average family in the index spends on various items of consumption. The amounts of commodities consumed and their pricing are taken into account when conducting such an investigation. The consumption pattern can thus be easily ascertained. It is necessary that the family budget enquiry amongst the class of people to whom the index series is applicable should be conducted during the base period.

**(3) Deciding on the items.**
The items on which the money is spent are classified into certain well–accepted groups. One of the choicest and most frequently used classification is –
(i) Food
(ii) Clothing
(iii) Fuel and lighting
(iv) House Rent
(v) Miscellaneous

**(4) Obtaining price quotations.**
The collection of retail pricing is an important but time-consuming and complex task.

This is because such prices might vary from place to place, shop to shop, and person to person. Price quotes should be sought from the areas where the persons in question live or shop. Some of the principles that should be followed while collecting retail price data for the purposes of building cost of living indices are given below.:

a. The retail prices should relate to a fixed list of items and for each item., the quality should be fixed by means of suitable specifications.
b. Retail prices should be those actually charged to consumers for cash sales.
c. Discount should be taken into account if it is automatically given to all customers.
d. In a period of price control or rationing, where illegal prices are charged openly, such prices should be taken into account along with the controlled prices.

**(5) Working on CPI.**

After collecting quotations from all the retail outlets, an average price for each of the index goods must be calculated. Such averages are calculated initially for the index's base period and then monthly if the index is maintained on a monthly basis. The method used to average the quotations should produce impartial estimates of average prices paid by the group as a whole. This, of course, will be determined by the method used to select retail shops as well as the scope of the index. Prices or their relatives must be weighted in order to be converted into index numbers. The need for weighting arises because the relative importance of various items for different classes of people is not the same. For this reason, the cost-of-living index is always a weighted index.

**Aggregate Expenditure Method**

We use the quantities of various commodities consumed by a certain segment of the population in the base year as weights in this method. The total spend for each commodity is then computed for each year. To do so, multiply the current year's price by the quantity or weight of the base year and add these products. Similarly, we must compute the total expenditure for each commodity in the base year.

Thus, in order to calculate the index numbers, we have to divide the total expenditure of the current year by the total expenditure of the base year and multiply the resulting figure by 100.

Consumer Price Index = $\frac{\sum p1}{\sum p0} \times 100$

Here,

$p_1$ = prices of the current year

$p_0$ = prices of the base year.

**Illustration 6.2.5**

From the following data compute the consumer price index for the year 2015 with reference to the base year 2012.

| Item: | Food | Clothing | Rent | Fuel | Sundries |
|---|---|---|---|---|---|
| Weight: | 8 | 3 | 2 | 2 | 3 |
| Percentage increase in price over 2012: | 60 | 40 | 20 | 30 | 40 |

**Solution**

**Computation of consumer price index for 2015**

| Items | Weight V | % increase in price | 100+ % increase I | IV |
|---|---|---|---|---|
| Food | 8 | 60 | 160 | 1280 |
| Clothing | 3 | 40 | 140 | 420 |
| Rent | 2 | 20 | 120 | 240 |
| Fuel | 2 | 30 | 130 | 260 |
| Sundries | 3 | 40 | 140 | 420 |
| | 18 | | | 2,620 |

Cost of living index $= \frac{\Sigma IV}{\Sigma V} = \frac{2620}{18} = 145.55$

**Family Budget Method**

Under this method, we study the family budgets of a large number of people and estimate the aggregate expenditure of the average family for various items. These values are used as weights. We then convert the current year's prices into price relatives on the basis of the base year's prices. We then multiply these price relatives by the respective values of the commodities of the base year. Now, we need to divide the total of these products by the sum of the weights.

**Fixed Base Index Number**

When index numbers for a number of years are computed serially on the basis of a fixed base year's data, it is a case of fixed base index numbers.

Such index numbers are otherwise called the price relatives. The base year thus fixed may be the remote most past year, any middle year, any recent past year, average of some years, or average of all the years given. The formula for computing such an index number is

$$P_{R1} = \frac{P1}{P0} \times 100$$

Where, $P_{R1}$ = Price relative of the current year
$P_1$ = Price of the current year.
$P_0$ = Price of the fixed base year.

**Illustration 6.2.6**

**(Fixed Base)**

From the following data relating to the average prices of a commodity compute the index numbers for each of the seven years taking 2010 as the base year.

| Year: | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|---|---|---|---|---|---|---|---|
| Average Prices (₹): | 50 | 60 | 55 | 65 | 75 | 80 | 70 |

**Solution**

**Calculation of fixed base index taking 2010 as base year.**

| Year | Prices (₹.) | Indices |
|---|---|---|
| 2010 | 50 | $\frac{50}{50} \times 100 = 100$ |
| 2011 | 60 | $\frac{60}{50} \times 100 = 120$ |
| 2012 | 55 | $\frac{55}{50} \times 100 = 110$ |
| 2013 | 65 | $\frac{65}{50} \times 100 = 130$ |
| 2014 | 75 | $\frac{75}{50} \times 100 = 150$ |
| 2015 | 80 | $\frac{80}{50} \times 100 = 160$ |
| 2016 | 70 | $\frac{70}{50} \times 100 = 140$ |

**Chain Index Numbers**

Under this method, firstly we express the figures for each year as a percentage of the preceding year. These are known as Link Relatives. We then need to chain them together by successive multiplication to form a chain index. Thus, unlike fixed base methods, in this method, the base year changes every year. Hence, for the year 2001, it will be 2000, for 2002 it will be 2001, and so on.

**Steps in the construction of Chain Index Numbers**

1. Calculate the link relatives by expressing the figures as the percentage of the preceding year. Link Relatives of current year = (price of current year/ price of previous year) × 100
2. Calculate the chain index by applying the following formula:
   Chain Index = (Current year relative ×Previous year link relative) / 100

**Illustration 6.2.7**

**(Chain Base)**

From the following data of the prices of certain goods, construct chain base index numbers.

| Year: | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|---|---|---|---|---|---|---|---|
| Prices in .₹: | 75 | 50 | 65 | 60 | 72 | 70 | 69 |

**Solution**

**Construction of chain base index**

| Year | Price in ₹ | Link relatives | Chain indices 2010=100 |
|---|---|---|---|
| 2010 | 75 | 100.00 | 100.00 |
| 2011 | 50 | $\frac{50}{75}\times 100 = 66.67$ | $\frac{66.67\times 100}{100} = 66.67$ |
| 2012 | 65 | $\frac{65}{50}\times 100 = 130.00$ | $\frac{130\times 66.67}{100} = 86.67$ |
| 2013 | 60 | $\frac{60}{65}\times 100 = 92.30$ | $\frac{92.30\times 86.67}{100} = 80.00$ |
| 2014 | 72 | $\frac{72}{60}\times 100 = 120.00$ | $\frac{120\times 80}{100} = 96.00$ |
| 2015 | 70 | $\frac{70}{72}\times 100 = 97.22$ | $\frac{97.22\times 96}{100} = 93.33$ |
| 2016 | 69 | $\frac{69}{70}\times 100 = 98.57$ | $\frac{98.57\times 93.33}{100} = 92.00$ |

**Difference between fixed base and chain base methods**

| Fixed base | Chain base |
|---|---|
| Base year is fixed | Base period is not fixed |
| Base period is arbitrarily chosen | Any period immediately preceding the current year is taken as base. |
| It is easily understood by a common man | It is not easily understood by a common man |
| Calculation is easy | Calculation is not easy |
| It is very rigid not permitting the accommodation of new items | It is quite flexible and so it easily permits the inclusion of some items |
| It does not permit frequent alteration of the weights of different items | It permits frequent adjustments of the weight of different items |
| It does not facilitate comparison between two adjacent period | It facilitates comparison between two adjacent periods |
| It is heavily affected by seasonal variation | It is least affected by seasonal variations |
| It does not create any problem in calculating the indices of the subsequent period when data of any year is missing | If data f any year is missing, it creates problems in the calculation of index numbers |
| It is suitable for long period | It is suitable for short period |
| An error in the calculation of any year does not vitiate the calculation of the balance of years | An error in the calculation of any one year will vitiate the calculation of the balance year |

## Recap

- Price index number - a statistical measure used to track and quantify changes in the average prices
- Price index number - serves as a vital tool for assessing inflation or deflation within an economy.
- Index provides a percentage or ratio that reflects the overall price level's variation.
- Wholesale Price Index Number- is a key indicator for measuring the dynamic movement of wholesale pricing
- Retail Price Index - measures changes in a sector's or community's price level
- Consumer price index - measures changes in the cost of goods and services purchased by households.

## Objective Questions

1. Which index is used to track changes in the cost of living for consumers?
2. What is the formula for calculating a price index number?
3. Which index is primarily concerned with tracking price fluctuations in manufacturing and industry?
4. What does WPI stand for?
5. Which type of index assigns weights to various items explicitly based on quantities or values?
6. What is the primary purpose of a Consumer Price Index (CPI)?
7. Which method of constructing index numbers uses quantities of various commodities consumed in the base year as weights?

## Answers

1. Consumer Price Index (CPI)
2. (Current year price / Base year price) x 100
3. Wholesale Price Index (WPI)
4. Wholesale Price Index (WPI)
5. Weighted Index Numbers
6. Assessing inflation or deflation
7. Aggregate Expenditure Method

## Self-Assessment Questions

1. How do price index numbers impact economic analysis, monetary policy, and salary adjustments?
2. What are the key components of the Consumer Price Index (CPI) basket of goods?
3. What is the difference between Simple Aggregate Method and Simple Average Relatives Method in constructing index numbers?
4. When is the Simple Aggregate Method considered less reliable, and what is its limitation?
5. What are weighted index numbers, and why are they used in constructing price indices?
6. List the types of weighted index numbers, and briefly describe each one.
7. Compare and contrast Laspeyre's and Paasche's methods for constructing weighted index numbers.
8. What is Fisher's Ideal Method, and why is it considered an improvement over Laspeyre's and Paasche's methods?
9. Explain the importance of the Marshall-Edgeworth Index (MEI) in constructing price indices and its relationship with Fisher's ideal index.

## Assignments

1. Describe the concept of price index numbers, their purpose, and how they are used to measure changes in the cost of living.
2. Describe the Wholesale Price Index (WPI), its significance in economic analysis, and provide examples of how it is used by businesses and policymakers.
3. Explain the Retail Price Index, its role in measuring changes in consumer prices, and how it differs from the Wholesale Price Index.
4. Discuss the Simple Aggregate Method for calculating price index numbers, including its formula and limitations. Provide an example of its application.
5. Explain the Simple Average Relatives Method for constructing price index numbers, its advantages, and disadvantages. Provide a practical example.
6. Define Weighted Index Numbers and discuss why they are used in price index calculations. Provide examples of different weighting methods and their significance.
7. Describe the Consumer Price Index (CPI), its purpose in measuring changes in the cost of living, and how it is calculated. Explain the uses of CPI in economic analysis.

## Suggested Readings


1. C.B Gupta and Vijay Gupta (2004) *An Introduction to Statistical Methods*. Vikas Publishing House.
2. S.P. Gupta. *Statistical Methods*. Sultan Chand and Sons, New Delhi.
3. Frederick E Croxton, Dudley J Cowden and Sidney Klein. *Applied General Statistics*. Prentice Hall India.
4. Naval Bajpai. *Business statistics*. Pearson Educational Publications
5. Dr. S M Shukla and Dr. Sahai. *Principles of statistics*. Sahitya Bhavan Publication, Delhi

# Unit - 3

# Quantity and Value Index Number

## Learning Outcomes

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

- ✓ familiarise with the concept of quantity index
- ✓ differentiate between the time reversal and factor reversal tests
- ✓ assess the importance of the unit test, time reversal test, and factor reversal test in checking the consistency of index numbers

## Prerequisites

In a busy city, there was a young shop owner named Rahul who sold gadgets like phones, laptops, and accessories. Over time, Rahul noticed that both the number of products he was selling and the prices were changing. He wanted to figure out how these changes were affecting his business. But before he could start, he realized he needed to focus on two things: how many items he sold and how much money he made from selling them. To do this, he needed to follow a few important steps.

“First, I need to decide what to track,” Rahul thought. “I can’t track everything.” He decided to focus on his best-selling items: smartphones, laptops, and accessories. For each of these, he needed to know two things: how many he sold and at what price. He knew he had to gather accurate sales records. “I’ll check my sales data for the last six months to see how many items I sold and how much I charged for them,” he said.

Next, Rahul realized he needed a starting point to compare everything. “I’ll use last year’s sales as the base year,” he decided. This would give him a normal year to compare against.

Finally, Rahul thought about how to calculate the changes. “I’ll calculate how much the number of items sold has changed and how much the total money I made has changed, using percentages,” he said. With this plan in place, Rahul could better understand how his sales were changing and make smarter choices for his store in the future.

## Keywords

Quantity index, Value index, Unit test, Time reversal test, Factor reversal test

## Discussion

### 6.3.1 Quantity index number

Quantity index number is a classification of index numbers which measures the changes in the quantity or volume of a particular variable within a specified timeframe. A base period is selected as reference and the current or subsequent periods are compared to this base period. They are useful in studying the level of physical output in an economy.

Quantity index number = $\frac{q_1}{q_0} \times 100$

Where,

$q_1$ is the quantity of the current year,

$q_0$ is the quantity of the base year

The resulting index number indicates how the quantity of the variable has changed relative to the base period. If the index number is greater than 100, it suggests an increase in quantity compared to the base period, while an index number less than 100 indicates a decrease.

For example, if you want to study how smartphone production has changed from the year 2020 to 2022. Suppose 2020 is the base year with 100000 smartphones produced, giving it an index of 100. In 2021, when 120000 smartphones were produced, the index is 120, indicating a 20% increase from the base year. Conversely, in 2022, with 90000 smartphones produced, the index is 90, signaling a 10% decrease from the base year. These index numbers simplify comparisons, where 100 represents the base year, values above 100 signify growth, and values below 100 indicate a decline. This method facilitates the analysis of trends in various data sets, such as economic indicators like the Consumer Price Index, enabling researchers to monitor changes effectively.

Similar to price index, quantity index can also be calculated using the following methods:

a) Laspeyre's quantity index- In this approach, we assign the base price as the weight, and we solely consider the price of the base year, excluding the prices of the current year.

$q_{01}$ (L) = $\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times 100$

b) Paashe's quantity index- In this context, the weight is determined by using the current year's price ($p_1$) for the commodity.

$q_{01}$ (P) = $\frac{\Sigma q_1 p_1}{\Sigma q_0 p_1} \times 100$

c) Fisher's quantity index- the formula to compute Fisher's quantity index number is:

$q_{01}$ (F) = $\sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}} \times 100$

## 6.3.2 Value index number

Value index numbers are not common like price index numbers and quantity index numbers. Value index number measures the changes in the level of value of items consumed during the year under study with reference to the level of value of item consumed in the base year. The value of an item is the price multiplied by quantity.

Value index number ($V_{01}$) = $\frac{\text{Total value of items consumed in current year}}{\text{Total value of items consumed in base year}} \times 100$

$= \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} \times 100$

**Illustration 6.3.1**

From the following data, calculate Laspeyre's quantity index number.

| Commodity | Price | | Quantity | |
|---|---|---|---|---|
| | **2022** | **2023** | **2022** | **2023** |
| A | 20 | 22 | 16 | 14 |
| B | 16 | 22 | 12 | 14 |
| C | 18 | 25 | 12 | 10 |
| D | 20 | 26 | 8 | 12 |

**Solution**

Construction of quantity index number

| Commodity | Price | | Quantity | | $q_1p_0$ | $q_0p_0$ |
|---|---|---|---|---|---|---|
| | $p_0$ | $p_1$ | $q_0$ | $q_1$ | | |
| A | 20 | 22 | 16 | 14 | 280 | 320 |
| B | 16 | 22 | 12 | 14 | 224 | 192 |
| C | 18 | 25 | 12 | 10 | 180 | 216 |
| D | 20 | 26 | 8 | 12 | 240 | 160 |
| | | | | | $\Sigma q_1p_0$ **924** | $\Sigma q_0p_0$ **888** |

Laspeyre's index $q_{01}$ (L) $= \frac{\Sigma q_1p_0}{\Sigma q_0p_0} \times 100$

$= \frac{924}{888} \times 100 = 104.054$

**Illustration 6.3.2**

From the following data, calculate Paashe's quantity index number.

| Commodity | Price | | Quantity | |
|---|---|---|---|---|
| | **2022** | **2023** | **2022** | **2023** |
| A | 22 | 26 | 16 | 14 |
| B | 18 | 26 | 12 | 16 |
| C | 20 | 25 | 14 | 10 |
| D | 29 | 33 | 8 | 12 |

**Solution**

Construction of quantity index number

| Commodity | Price | | Quantity | | $q_1p_1$ | $q_0p_1$ |
|---|---|---|---|---|---|---|
| | $p_0$ | $p_1$ | $q_0$ | $q_1$ | | |
| A | 22 | 26 | 16 | 14 | 364 | 416 |
| B | 18 | 26 | 12 | 16 | 416 | 312 |
| C | 20 | 25 | 14 | 10 | 250 | 350 |
| D | 29 | 33 | 8 | 12 | 396 | 264 |
| | | | | | $\Sigma q_1p_1$ **1426** | $\Sigma q_0p_1$ **1342** |

Paashe's quantity index $q_{01}$ (P) = $\frac{\Sigma q_1 p_1}{\Sigma q_0 p_1} \times 100$

$= \frac{1426}{1342} \times 100$

= 106.259

**Illustration 6.3.3**

From the following data, calculate Fisher's quantity index number.

| Commodity | Price | | Quantity | |
|---|---|---|---|---|
| | 2022 | 2023 | 2022 | 2023 |
| A | 20 | 22 | 6 | 4 |
| B | 20 | 26 | 2 | 6 |
| C | 25 | 27 | 4 | 10 |
| D | 29 | 33 | 8 | 2 |

**Solution**

Construction of quantity index number

| Commodity | Price | | Quantity | | $q_1p_1$ | $q_0p_1$ | $q_0p_0$ | $q_1p_0$ |
|---|---|---|---|---|---|---|---|---|
| | $p_0$ | $p_1$ | $q_0$ | $q_1$ | | | | |
| A | 20 | 22 | 6 | 4 | 88 | 132 | 120 | 80 |
| B | 20 | 26 | 2 | 6 | 156 | 52 | 40 | 120 |
| C | 25 | 27 | 4 | 10 | 270 | 108 | 100 | 250 |
| D | 29 | 33 | 8 | 2 | 66 | 264 | 232 | 58 |
| | | | | | $\Sigma q_1p_1$ **580** | $\Sigma q_0p_1$ **556** | $\Sigma q_0p_0$ **492** | $\Sigma q_1p_0$ **508** |

Fisher's quantity index $q_{01}$ (F) = $\sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}} \times 100$

$= \sqrt{\frac{508}{492} \times \frac{580}{556}} \times 100$

$= \sqrt{1.0325 \times 1.0431} \times 100$

$= \sqrt{1.077} \times 100$

= 103.778

**Illustration 6.3.4**

From the following data, calculate value index number.

| Commodity | 2022 | | 2023 | |
|---|---|---|---|---|
| | **Price** | **Expenditure** | **Price** | **Expenditure** |
| A | 5 | 30 | 15 | 75 |
| B | 9 | 18 | 20 | 60 |
| C | 9 | 54 | 10 | 80 |
| D | 6 | 30 | 12 | 48 |

**Solution**

To compute the quantity of commodities A, B, C, and D, we have to divide expenditure by price of each commodity.

Quantity of commodity A in 2022 = $\frac{30}{5} = 6$

Quantity of commodity A in 2023 = $\frac{75}{15} = 5$

Construction of value index number

| Commodity | Price | | Quantity | | $p_1q_1$ | $p_0q_0$ |
|---|---|---|---|---|---|---|
| | $p_0$ | $q_0$ | $p_1$ | $q_1$ | | |
| A | 5 | 6 | 15 | 5 | 75 | 30 |
| B | 9 | 2 | 20 | 3 | 60 | 18 |
| C | 9 | 6 | 10 | 8 | 80 | 54 |
| D | 6 | 5 | 12 | 4 | 48 | 30 |
| | | | | | $\sum p1q1$ **263** | $\sum p0q0$ **132** |

Value index number = $\frac{\sum p_1 q_1}{\sum p_0 q_0} \times 100$

$= \frac{263}{132} \times 100 = 199.242$

## 6.3.3 Tests of index number

Consistency is an important property of index numbers, which are used to measure changes in the relative value of a set of items over time or across different groups. There are several mathematical tests and criteria that can be used to assess the consistency of index numbers. These tests help ensure that the index numbers accurately reflect the underlying data and are suitable for making meaningful comparisons. Here are some of the key mathematical tests of consistency for index numbers:

i. Unit test
ii. Time reversal test
iii. Factor reversal test

### 6.3.3.1 Unit tests

The unit test is a mathematical test of consistency of index number that ensures that the index number is independent of the units in which prices and quantities are expressed. This means that the index number should not change if the prices and quantities are expressed in different units, such as kilograms or pounds, or rupees or dollars. This makes the index number more reliable as a measure of changes in prices and quantities.

### 6.3.3.2 Time reversal tests

This test has been put forth by Prof. Irving Fisher, who proposes that a formula of index number should be such that it turns the value of the index number to its reciprocal when the time subscripts of the formula are reversed. This method is a device to determine if a method will work both ways in time 'backward and forward'.

The time reversal test is a mathematical test used to check the consistency of an index number when comparing two periods in opposite directions. In other words, it examines whether the index value remains the same when you reverse the direction of comparison between two time periods. If the time reversal test is satisfied, it means that the index is consistent when comparing Period, A to Period B and Period B to Period A. The test helps to ensure that the index number accurately reflects the relative changes between the two periods, regardless of the order in which they are compared.

For example, if the price of a commodity has increased to ₹20 per kilogram in 2022, as compared to ₹10 per kilogram in 2021, we would say that the price in 2022 is 200 per cent of the price in 2021 and the price in 2021 is 50 per cent of the 2022 price. Now these two figures are reciprocal of one another and their product (2.00 x 0.50) is equal to unity. If the method does not work both ways, i.e., if the index number for two years secured by the same method but with basis reversed are not reciprocal of each other there is an inherent bias in the method.

Algebraically, the test may be expressed as

$P_{01} \times P_{10} = 1$

Where $P_{01}$ stands for index for the current year on the base year omitting the factor 100, (i.e., for price change in current year as compared with base year) and $P_{10}$ stands for index for the base year on the current year without the factor 100 (i.e., for the price change in base year compared with current year).

According to Fisher, "the test is that the formula for calculating an index number should be such that it will give the same ratio between one point of comparison and the other, no matter which of the two is taken as base," or putting it another way, "the index number reckoned forward should be the reciprocal of that reckoned backward."

That Fisher's ideal index satisfies the 'Time Reversal Test' can also be seen from the following illustration:

$$P_{01} = \sqrt{\frac{\sum P_1 q_0}{\sum P_0 q_0} \times \frac{\sum P_1 q_1}{\sum P_0 q_1}}$$

Changing current year to base

$$P_{10} = \sqrt{\frac{\sum P_0 q_0}{\sum P_1 q_0} \times \frac{\sum P_0 q_1}{\sum P_1 q_1}}$$

Time Reversal test is: $P_{01} \times P_{10} = 1$

$$\text{Now } P_{01} \times P_{10} = \sqrt{\frac{\sum P_1 q_0}{\sum P_0 q_0} \times \frac{\sum P_1 q_1}{\sum P_0 q_1} \times \frac{\sum P_0 q_0}{\sum P_1 q_0} \times \frac{\sum P_0 q_1}{\sum P_1 q_1}}$$

$$= 1$$

Thus, we see that the indices prepared according to Fisher's ideal formula satisfy the Time Reversal Test.

### 6.3.3.3 Factor reversal test

This test has also been put forth by Prof. Irving Fisher, who proposes that a formula of index number should be such that it permits the interchange of the price, and the quantity factors without giving inconsistent result i.e., the two results multiplied together should give the true ratio in as much as the product of price and quantity is the value of a thing.

According to Fisher "Just as our formula should permit the interchange of the two times without giving inconsistent results so it ought to permit interchanging the prices and quantities without giving inconsistent result, i.e., the two results multiplied together should give the true ratio"

In simple words the test is satisfied if the product of the price index and the quantity index is equal to the ratio of the aggregate value (quantity x price) in the current year to the aggregate value in the base year.

Algebraically: $p_{01} \times q_{01} = \frac{\sum P_1 q_1}{\sum P_0 q_0}$

Where $P_{01}$ stands for the price change for the current year over the base year, $q_{01}$ stands for the quantity change for the current year over the base year, $\sum P_1q_1$ stands for the total value in the current year, and $\sum P_0q_0$ stands for the total value in the base year.

That Fisher's ideal index satisfies this test can be seen from the following example:

| Commodity | 2021 | | 2022 | | $p_0q_0$ | $p_1q_0$ | $p_0q_1$ | $p_1q_1$ |
|---|---|---|---|---|---|---|---|---|
| | $p_0$ | $q_0$ | $p_1$ | $q_1$ | | | | |
| A | 5 | 12 | 7 | 15 | 60 | 84 | 75 | 105 |
| B | 7 | 10 | 9 | 12 | 70 | 90 | 84 | 108 |
| C | 9 | 8 | 10 | 9 | 72 | 80 | 81 | 90 |
| D | 10 | 5 | 13 | 6 | 50 | 65 | 60 | 78 |
| E | 11 | 3 | 15 | 4 | 33 | 45 | 44 | 60 |
| | | | | | **285** | **364** | **344** | **441** |

Factor Reversal Test is satisfied, if

$$p_{01} \times q_{01} = \frac{\sum P_1q_1}{\sum P_0q_0}$$

Where $p_{01}$ stands for the price change for the current year over the base year, and $q_{10}$ stands for the changes for the current year over the base year.

Now according to Fisher's ideal index number formula:

$$P_{01} = \sqrt{\frac{\sum P_1q_0}{\sum P_0q_0} \times \frac{\sum P_1q_1}{\sum P_0q_1}}$$

and $$q_{01} = \sqrt{\frac{\sum P_0q_1}{\sum P_0q_0} \times \frac{\sum P_1q_1}{\sum P_1q_0}}$$

Hence,

$$P_{01} \text{ x } q_{01} = \sqrt{\frac{\sum P_1q_0}{\sum P_0q_0} \times \frac{\sum P_1q_1}{\sum P_0q_1} \times \frac{\sum P_0q_1}{\sum P_0q_0} \times \frac{\sum P_1q_1}{\sum P_1q_0}}$$

As per the aforementioned table, we get

$$P_{01} \times q_{01} = \sqrt{\frac{364}{285} \times \frac{441}{344} \times \frac{344}{285} \times \frac{441}{364}}$$

$$= \sqrt{\frac{441}{285} \times \frac{441}{285}} = \frac{441}{285}$$

Now $\frac{\sum P_1 q_1}{\sum P_0 q_0}$ is also equal to $\frac{441}{285}$

Thus, it is proved that Fisher's ideal formula for index number satisfies the 'Factor Reversal Test.'

**Illustration 6.3.5**

From the following data calculate Fisher's ideal index number and see whether it satisfies both Time Reversal and Factor Reversal Tests.

| **Commodity** | **2020** | | **2022** | |
|---|---|---|---|---|
| | **Price** | **Expenditure** | **Price** | **Expenditure** |
| M | 3 | 21 | 5 | 40 |
| N | 5 | 30 | 8 | 24 |
| O | 7 | 56 | 9 | 45 |
| P | 6 | 18 | 7 | 28 |
| Q | 10 | 20 | 12 | 24 |

**Solution**

We are given price and expenditure of each commodity. As we are not given quantity, it should be obtained by dividing expenditure by price of each quantity.

Thus, the quantity of commodity M in 2020 = $\frac{21}{3} = 7$

and the quantity of commodity M in 2022 = $\frac{40}{5} = 8$

Calculation of Fisher's ideal index number

| Commodity | 2020 | | 2022 | | $p_0q_0$ | $p_1q_0$ | $p_0q_1$ | $p_1q_1$ |
|---|---|---|---|---|---|---|---|---|
| | $p_0$ | $q_0$ | $p_1$ | $q_1$ | | | | |
| M | 3 | 7 | 5 | 8 | 21 | 35 | 24 | 40 |
| N | 5 | 6 | 8 | 3 | 30 | 48 | 15 | 24 |
| O | 7 | 8 | 9 | 5 | 56 | 72 | 35 | 45 |
| P | 6 | 3 | 7 | 4 | 18 | 21 | 42 | 28 |
| Q | 10 | 2 | 12 | 2 | 20 | 24 | 20 | 24 |
| | | | | | **145** | **200** | **136** | **161** |

Fisher's ideal index number $= \sqrt{\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1}} \times 100$

$$= \sqrt{\frac{200}{145} \times \frac{161}{136}} \times 100$$

$$= \sqrt{1.379 \times 1.183} \times 100$$

$$= 127.72$$

Time Reversal Test is satisfied if, $= P_{01} \times P_{10} = 1$

Now $P_{01} \text{ x } P_{10} = \sqrt{\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1} \times \frac{\Sigma P_0 q_0}{\Sigma P_1 q_0} \times \frac{\Sigma P_0 q_1}{\Sigma P_1 q_1}}$

$$= \sqrt{\frac{200}{145} \times \frac{161}{136} \times \frac{145}{200} \times \frac{136}{161}}$$

$$= 1$$

Since the answer is 1, it is evident that Fisher's ideal index number satisfies the Time Reversal Test.

Factor Reversal Test is satisfied, if $P_{01} \times q_{01} = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_0}$

i.e., $P_{01} \times q_{01} = \sqrt{\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1} \times \frac{\Sigma P_0 q_1}{\Sigma P_0 q_0} \times \frac{\Sigma P_1 q_1}{\Sigma P_1 q_0}} = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_0}$

$$= \sqrt{\frac{200}{145} \times \frac{161}{136} \times \frac{136}{145} \times \frac{161}{200}}$$

$$= \sqrt{\frac{161}{145} \times \frac{161}{145}} = \frac{161}{145}$$

Now $\frac{\Sigma P_1 q_1}{\Sigma P_0 q_0}$ is also equal to $\frac{161}{145}$

Thus, it is proved that Fisher's ideal formula for index number satisfies the 'Factor Reversal Test.'

## Recap

- Quantity index number- Change in quantity over a period
- Value index number- Change in value over a period
- Value = price × quantity
- Quantity index number = $\frac{q_1}{q_0} \times 100$
- Value index number = $\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} \times 100$
- $q_{01}$ (L) = $\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times 100$
- $q_{01}$ (P) = $\frac{\Sigma q_1 p_1}{\Sigma q_0 p_1} \times 100$
- $q_{01}$ (F) = $\sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}} \times 100$
- Unit test - Checking whether the index reflects changes accurately when measured in the same units as the original data.
- Time Reversal Test - Ensuring it is not sensitive to the direction of time.
- Factor Reversal Test- Evaluates how well an index responds to changes in the weights or factors used in its construction.
- Ideal index number - Fisher's index number is said to be an ideal index number because it satisfies both time and factor Reversal test.

## Objective Questions

1. What is the primary purpose of a quantity index in economic analysis?
2. What is the primary purpose of a value index in economic analysis?
3. In Quantity index formula, what does $q_0$ represent?
4. How is value calculated in value index number?
5. What is the purpose of the unit test in index numbers?
6. What is the primary purpose of the time reversal test in index number construction?
7. What does the factor reversal test examine?
8. If an index passes both time reversal and factor reversal tests, what does it imply?

## Answers

1. To measure changes in the quantity of goods and services
2. To measure changes in both price and quantity simultaneously
3. The base period quantities of goods and services
4. Value = price x quantity
5. To check accuracy for a single unit
6. To check for time-related bias
7. Quantity-related bias
8. It is less biased and more reliable.

## Self-Assessment Questions

1. What does quantity index number measures?
2. What are the three methods of quantity index numbers?
3. What do you mean by value index number?
4. Compare and contrast the unit test, time reversal test, and factor reversal test in terms of their objectives and applications in index number construction.
5. Why is Fisher's index number considered an ideal index number?
6. What is the importance of the time reversal test?
7. What is the importance of the factor reversal test?
8. What is unit test?
9. Why should we check the consistency of an index number?
10. How can we check the consistency of an index number?

## Assignments

1. Compute quantity index number for the year 2023 with 2020 as the base year using:
   a. Laspeyre's method
   b. Paasche's method
   c. Fisher's method

| Commodity | Quantity | | Value | |
|---|---|---|---|---|
| | 2020 | 2023 | 2020 | 2023 |
| A | 100 | 150 | 500 | 9000 |
| B | 80 | 100 | 320 | 300 |
| C | 60 | 72 | 150 | 360 |
| D | 30 | 33 | 360 | 297 |

(Answer: $q_{01}$ (L)- 129.744, $q_{01}$ (P)- 131.019, $q_{01}$ (F)- 130.395)

2. The price and quantity demanded for three products are given below:

| Product | Price (Year 1) | Price (Year 2) | Quantity (Year 1) | Quantity (Year 2) |
|---|---|---|---|---|
| A | 10 | 12 | 100 | 110 |
| B | 20 | 18 | 50 | 55 |
| C | 5 | 6 | 80 | 85 |

Calculate the Fisher price index for these two years and check whether it satisfies the factor reversal test.

(Answer $P_{01}$ (F) =107.464)

3. The price and quantity demanded for three commodities are given below:

| Product | Price 2021 | Price 2022 | Quantity 2021 | Quantity 2022 |
|---|---|---|---|---|
| X | 11 | 15 | 75 | 107 |
| Y | 18 | 22 | 65 | 56 |
| Z | 7 | 13 | 81 | 95 |

Calculate the Fisher price index for these two years and check whether it satisfies the time reversal test.

(Answer $P_{01}$ (F) =141.84)

4. From the following data calculate Fisher's ideal index number and see whether it satisfies both time reversal and factor reversal tests.

| Commodity | 2021 | | 2022 | |
|---|---|---|---|---|
| | Price | Expenditure | Price | Expenditure |
| D | 12 | 24 | 15 | 45 |
| E | 9 | 36 | 12 | 60 |
| F | 13 | 65 | 16 | 32 |
| G | 8 | 16 | 14 | 98 |
| H | 10 | 20 | 12 | 72 |

(Answer $P_{01}$ (F) =134.002)

## Suggested Readings


1. Box and Tiao. *Time Series Analysis*, Forecasting, and Control, Holden Day.
2. C.B Gupta and Vijay Gupta (2004) *An Introduction to Statistical Methods*. Vikas Publishing House.
3. S.P. Gupta. *Statistical Methods*. Sultan Chand and Sons, New Delhi.
4. Frederick E Croxton, Dudley J Cowden and Sidney Klein. *Applied General Statistics*. Prentice Hall India.
5. Naval Bajpai. *Business statistics*. Pearson Educational Publications
6. Dr. S M Shukla and Dr. Sahai. *Principles of statistics*. Sahitya Bhavan Publication, Delhi

**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

# Business Mathematics and Statistics

COURSE CODE: SGB24CM102MC



ISBN 978-81-984516-0-6

Sreenarayanaguru Open University
Kollam, Kerala Pin- 691601, email: info@sgou.ac.in, www.sgou.ac.in Ph: +91 474 2966841