Your road to success

Your road to success

LEVEL 4 INTRODUCTION TO QUANTITATIVE METHODS

ABE • OFFICIAL STUDY GUIDE

abeuk.com

ISBN: 978-1-911550-13-6

First published in 2017 by ABE
5th Floor, CI Tower, St. Georges Square, New Malden, Surrey, KT3 4TE, UK
www.abeuk.com

All facts are correct at time of publication.

Author: Rajesh Raheja BSc MBA

Reviewer: Colin Linton MRes MBA PGCHE DipM DipFS FCIB FCIM FCIPS FCIEA FHEA FInstLM

Editorial and project management by Haremi Ltd.
Typesetting by York Publishing Solutions Pvt. Ltd., INDIA

Every effort has been made to trace all copyright holders, but if any have been inadvertently overlooked, the Publishers will be pleased to make the necessary arrangements at the first opportunity.

The rights of Rajesh Raheja BSc MBA to be identified as the author of this work have been asserted by her in accordance with the Copyright, Design and Patents Act 1998.

Acknowledgements
The publishers gratefully acknowledge permission to reproduce the following copyright material:
P23 Jat306/Shutterstock; p23 SFIO CRACHO/ Shutterstock; p24 Casper1774 Studio/Shutterstock; p27 kurhan/Shutterstock; p28 SARAVOOT LENG-IAM/ Shutterstock; p29 kurhan/Shutterstock; p30 kurhan/Shutterstock; p41 gabczi/ Shutterstock; p45 ChristianChan/Shutterstock; p47 Andrey_Popov/Shutterstock; p49 wutzkohphoto/Shutterstock; p50 Thanakit Jitkasem/Shutterstock; p54 Elena Shashkina/Shutterstock; p58 varuna/Shutterstock; p59 Artazum/ Shutterstock; p63 attaphong/Shutterstock; p64 sheff/Shutterstock; p65 como/ Shutterstock; p72 Olesia Bilkei/Shutterstock; p79 hxdbzxy/Shutterstock; p82 GLF Media/Shutterstock; p103 Anna-Mari West/Shutterstock; p108 Foxys Forest Manufacture/Shutterstock; p122 Viktor1/Shutterstock; p147 ndoeljindoel/Shutterstock; p151 Lamai Prasitsuwan/Shutterstock;

Contents

Using your study guide

Welcome to the study guide for the **Level 4 Introduction to Quantitative Methods**.

Below is an overview of the elements of learning and related key capabilities (taken from the published syllabus), designed to support learners to assess their own skillset in terms of employability and to develop their own personal development plans.

Element of learning	Key capabilities
Element 1 – Numeracy for business	Apply numerical techniques in context typical business situations *Problem posing and problem solving using arithmetic, numeracy, application of mathematical formulas*
Element 2 – Algebraic methods	Use algebraic methods to express relationship between variables and find their values. *Problem posing and problem solving using algebra, mapping and presenting relationships between variables* Plot and interpret mathematical graphs *Visual presentation and analysis of information*
Element 3 – Business statistics	Assessing data with the application of statistical techniques to gain insight into real world business environment *Planning research, creative data collection, analysing and interpreting data, presenting information*
Element 4 – Statistical tools and data analysis	Employing statistical techniques to interpret data and communicate quantitative results for managerial decision making *Application of statistical tools, data analysis and synthesis, presentation of information, creative interpretation for evaluation*

This study guide follows the order of the syllabus, which is the basis for your studies. Each chapter starts by listing the overarching syllabus learning outcomes covered and the assessment criteria.

L4 descriptor

Knowledge descriptor (the holder...)	Skills descriptor (the holder can...)
• Has practical, theoretical or technical knowledge and understanding of a subject or field of work to address problems that are well defined but complex and non-routine. • Can analyse, interpret and evaluate relevant information and ideas. • Is aware of the nature and approximate scope of the area of study or work. • Has an informed awareness of different perspectives or approaches within the area of study or work	• Identify, adapt and use appropriate cognitive and practical skills to inform actions, and address problems that are complex and non-routine while normally fairly well-defined. • Review the effectiveness and appropriateness of methods, actions and results.

The study guide includes a number of features to enhance your studies:

'Over to you': activities for you to complete, using the space provided.

Case studies: realistic business scenarios to reinforce and test your understanding of what you have read.

'Revision on the go': use your phone camera to capture these key pieces of learning, then save them on your phone to use as revision notes.

Examples: illustrating points made in the text to show how it works in practice.

Tables, graphs and charts: to bring data to life.

Reading list: identifying resources for further study, including Emerald articles (which will be available in your online student resources).

Source/quotation information to cast further light on the subject from industry sources.

Highlighted words throughout and **glossary terms** at the end of the book.

Note

Website addresses current as at June 2017.

Chapter 1

Numeracy for Business

Introduction

Numbers are an integral part of business management theory and practice. There is no economic activity that can take place without the use of numbers. Therefore, learning to communicate with the language of numbers is as important as learning to communicate with words. A simple example, that is relevant for almost everyone in today's world, is the ability to understand how banks calculate interest on savings and loans.

Learning outcome

On completing the chapter, you will be able to:

1 **Apply numeracy and quantitative techniques for use in day-to-day business activities. (Weighting 25%)**

Assessment criteria

1.1 Perform calculations on different types of numbers

1.2 Express numbers in various forms for making comparisons

1.3 Perform simple financial calculations to obtain values for taking business decisions

1.1 Performing calculations on different types of numbers

In this section, we will learn about different types of numbers and the rules that apply to them.

Application of rules of numeracy

Numbers can be classified into different types. The classification of numbers is given in Table 1.

Natural numbers	Natural numbers are all positive numbers starting from 1. Zero is not a natural number. The letter "n" denotes a natural number in arithmetic expressions. **Example:** 1,2,3,4…
Whole numbers	Whole numbers include all natural numbers and also zero. These are numbers with no decimal or fractional parts. These numbers never take negative values. **Example**: 0,1,2,3…
Integers	Integers include whole numbers and also negative numbers. These numbers do not have decimal or fractional parts and only represent whole units of positive and negative numbers. **Example**: …–4, –3, –2, –1, 0, 1, 2, 3, 4…
Fractions	Fractions are parts of whole numbers. These indicate that an integer is being divided by another integer. **Example**: $\frac{1}{2}, \frac{3}{4}, \frac{9}{11}$ …
Decimals	Decimals are numbers based on the decimal system or the 10s system. Decimals use a dot or a point to separate the ones place from the tenths place in a number. There are one or more digits to the right-hand side of the dot or point called the decimal point. **Example**: 32.65, 0.054, 1.2 …

Table 1: Types of numbers

REVISION on the go

> *Four* ***arithmetic operations*** *are:*
> *Addition, Subtraction, Multiplication and Division*

Just like a traditional language is built on certain grammatical principles, there are certain rules in application of numbers. It is important to follow these rules for correct numerical expression. When these rules are not followed, it results in meaningless and incorrect numerical communication.

The key rules to numeracy include:

- **BEDMAS** rule;
- rules for negative numbers;
- rules for fractions;
- rules for decimals.

BEDMAS rule

Numeric expressions can be both, simple and complex. Simple numerical expressions have just one arithmetic operation that may be addition, subtraction, multiplication or division. However, a complex numeric expression typically has a combination of different arithmetic operations. Therefore, to understand a complex numerical expression you need to apply BEDMAS, a rule that defines the order in which the arithmetic expression is simplified. BEDMAS is an acronym:

B – Brackets
E – Exponents
D – Division
M – Multiplication
A – Addition
S – Subtraction

This **order of operations** in the acronym simply implies that, according to the rules of numeracy, you would perform the operations division and multiplication only after opening brackets and simplifying exponents. However, both division and multiplication precede addition and subtraction.

You should understand that while division appears before multiplication, and addition appears before subtraction in the acronym BEDMAS, division and multiplication are considered higher level operations, and therefore, are applied before addition and subtraction. However, both multiplication and division are given equal importance, implying that you can apply either of these two operations before the other. Once you have completed these operations, only then can you apply addition and subtraction in any order because both these operations are given equal importance. If there are brackets within brackets then the innermost brackets need to be simplified and opened first.

Let's look at the application of BEDMAS rule in expressing $16 \div 2^2 + 6 + (4 \times 3 - [6 + 2 - 3])$ in its simplest form with an example.

Example

$16 \div 2^2 + 6 + (4 \times 3 - [6 + 2 - 3])$

$= 16 \div 2^2 + 6 + (4 \times 3 - 5)$

$= 16 \div 2^2 + 6 + 7$

$= 16 \div 4 + 6 + 7$

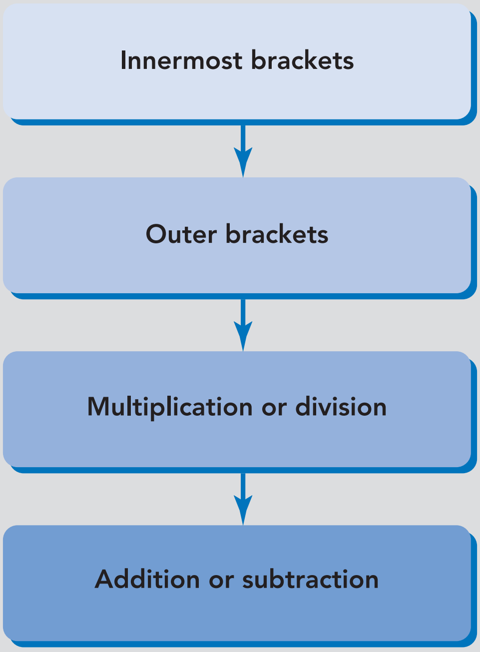
$= \frac{16}{4} + 6 + 7$

$= 4 + 6 + 7$

$= 17$

Note the order of operations: innermost brackets were simplified before other brackets, exponent, multiplication and division, and in the end, addition and subtraction.

Rules of numeracy



REVISION on the go

OVER TO YOU

Activity 1

Solve the following:

1. $(60 \times 3 - 5) + 25 \div 5$
2. $15 \div 3 \times (18 + 2 - [3 + 3^2])$
3. $88 + 88 \div (2 \times 11) \times 2$

Rules for negative numbers

Integers can be positive or negative numbers. A negative number is identified by the "–" sign as its prefix. A positive number may NOT have + as prefix. Arithmetic operations, when performed on two positive numbers always result in a positive number. However, when arithmetic operations involve either a positive number ("Positive") and a negative number ("Negative") or two negative numbers, the result is based on certain rules.

The rules of negative numbers are summarised in Table 2.

Rule	Examples
Positive + Negative = Either Positive or Negative number (depending on the bigger number)	12 – 13 = –1 13 – 12 = +1
Positive × Negative = Negative Number	5 × (–6) = –30
Negative × Negative = Positive Number	(–5) × (–6) = 30
Negative ÷ Positive or Positive ÷ Negative = Negative Number	(–12) ÷ 2 = –6 12 ÷ (–2) = –6
Negative ÷ Negative = Positive Number	(–12) ÷ (–2) = 6

Table 2: Rules of negative numbers

REVISION on the go

OVER TO YOU

Activity 2

When you multiply a positive number by another positive number, is the result positive or negative?

Numeracy rules for fractions

A fraction has a **numerator** and a **denominator**. Fractions can be of different types. These are explained in Table 3.

Fraction	Type	Example
An equivalent fraction	May look different but have the same value.	$\frac{1}{2} = \frac{8}{16}$
A proper fraction	In which the numerator is smaller than the denominator.	$\frac{5}{14}, \frac{17}{23}, \frac{25}{33}$
An improper fraction	In which the numerator is larger than the denominator.	$\frac{5}{4}, \frac{7}{3}, \frac{25}{13}$
A mixed fraction	Consists of a whole number alongside a fraction. The whole number always precedes the fraction.	$1\frac{1}{4}, 2\frac{2}{3}, 4\frac{5}{13}$

Table 3: Types of fractions

REVISION on the go

OVER TO YOU

Activity 3

Classify the following as proper, improper and mixed fractions:

1 $\frac{13}{4}$

2 $2\frac{1}{5}$

3 $\frac{3}{4}$

4 $\frac{11}{3}$

Certain rules must be followed for solving expressions that contain fractions. These rules are explained with examples in Table 4.

Rules for fractions	Examples
A fraction should always be simplified and should be expressed as the smallest value of the numerator and denominator. This involves cancelling out the numerator and denominator with a common whole number that divides both the numerator and denominator. Continue to cancel till the point when no more division is possible. Every cancelling step must include both numerator and denominator. Cancelling is possible while multiplying and dividing numbers. However, two separate fractions being added or subtracted cannot be cancelled against each other.	$\frac{5}{15} = \frac{\cancel{5}^1}{\cancel{15}^3} = \frac{1}{3}$ or $\frac{15}{75} = \frac{\cancel{15}^3}{\cancel{75}^{15}} = \frac{\cancel{3}^1}{\cancel{15}^3} = \frac{1}{5}$ Similarly, $\frac{\cancel{2}^1}{\cancel{25}^5} \times \frac{\cancel{15}^3}{\cancel{26}^{13}}$ $= \frac{1}{5} \times \frac{3}{13}$ $= \frac{1 \times 3}{5 \times 13} = \frac{3}{65}$ However, $\frac{1}{\cancel{4}^2} + \frac{1}{\cancel{2}^1} \neq \frac{1}{2} + \frac{1}{1}$
An improper fraction can be changed into a mixed fraction by dividing the numerator by the denominator. Any remainder is placed as a numerator over the denominator.	$\frac{25}{13}$ is changed into a mixed fraction by dividing 25 by 13. The resulting fraction is: $1\frac{1}{13}$

Rules for fractions	Examples
A mixed fraction can be converted into an improper fraction by: 1 multiplying the denominator with the whole number part of the mixed fraction; 2 adding the numerator to it. This will give the value of the numerator. The denominator stays the same.	$1\frac{12}{13} = (1 \times 13) + 12 = \frac{25}{13}$
Proper fractions can be added together or subtracted from each other. The steps are: 1 make the denominators equal, by finding a common denominator; 2 Then, find the numerators by simply adding or subtracting, as may be the case.	$\frac{1}{8} + \frac{5}{32} - \frac{3}{16}$ Since 32 is divisible by both 8 and 16, 32 is the common denominator and is taken as the denominator of the answer. We need to multiply 8 by 4 and 16 by 2 to get 32. To find the numerator we will multiply the numerator of the first fraction by 4 and the numerator of the third fraction by 2 and then simply perform the addition and subtraction. Therefore, $\frac{1}{8} + \frac{5}{32} - \frac{3}{16}$ $= \frac{(1 \times 4) + 5 - (3 \times 2)}{32}$ $= \frac{4 + 5 - 6}{32} = \frac{3}{32}$
Mixed fractions can be added and subtracted by separating whole numbers from fractions. Whole numbers are solved separately from fractions, and the sum of both is added together.	$3\frac{1}{8} + 1\frac{5}{32} - 2\frac{3}{16}$ $= (3 + 1 - 2) + \left(\frac{1}{8} + \frac{5}{32} - \frac{3}{16}\right)$ $= 2 + \frac{3}{32}$ $= 2\frac{3}{32}$
To multiply a fraction by a whole number, multiply the numerator by the whole number and leave the denominator the same.	$4 \times \frac{3}{7} = \frac{4 \times 3}{7} = \frac{12}{7} = 1\frac{5}{7}$
To divide a fraction by a whole number, multiply the denominator by the whole number and leave the numerator the same.	$\frac{3}{7} \div 4 = \frac{3}{7} \times \frac{1}{4} = \frac{3}{7 \times 4} = \frac{3}{28}$
To multiply a fraction by another fraction, multiply the numerators of the two fractions to get the numerator of the answer and the denominators of the two fractions to get the denominator of the answer.	$\frac{4}{5} \times \frac{3}{7} = \frac{4 \times 3}{5 \times 7} = \frac{12}{35}$

Rules for fractions	Examples
To divide a fraction by another fraction, take the reciprocal of the second fraction (divide 1 by it). Next multiply the first fraction with the reciprocal of the second fraction.	$\frac{4}{5} \div \frac{3}{7} = \frac{4}{5} \times \frac{7}{3} = \frac{28}{15} = 1\ \frac{13}{15}$
To divide a whole number by a fraction, take the reciprocal of the fraction. Next multiply the whole number by the reciprocal of the second fraction. Note that 6 is shown as being divided by 3.	$6 \div \frac{3}{7} = 6 \times \frac{7}{3} = \cancel{6}^{2} \times \frac{7}{\cancel{3}_{1}} = 14$

Table 4: Rules for calculating with fractions

Numeracy rules for decimals

There are rules for expressions that contain decimals. These are explained with examples in Table 5.

Rules for decimals	Example
While adding or subtracting decimals, the decimal points need to be aligned.	1.034 + 3.20 = 1.034 + 3.200 = 4.234
A decimal can be multiplied by 10 by simply moving the decimal point one place to the right, by 100 by moving the decimal point two places to the right and so on.	36.324 × 100 = 3632.4
A decimal can be divided by 10 by simply moving the decimal point one place to the left, by 100 by moving the decimal point two places to the left and so on.	36.324 ÷ 100 = 0.36324

Table 5: Rules for decimals

OVER TO YOU

Activity 4

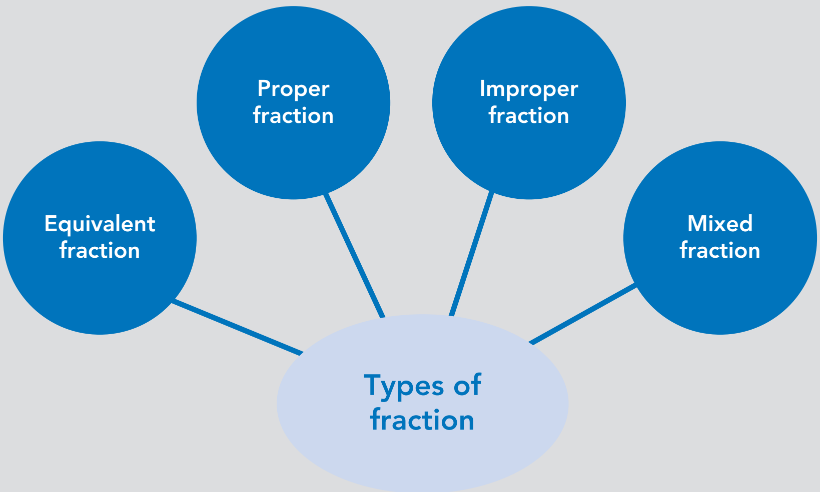
Solve the following:

1 $2\frac{1}{2} + \frac{8}{9} \times 2 - \left(1\frac{1}{2} - \frac{4}{5}\right)$

2 $\sqrt{\frac{9}{4}} \div \frac{3}{6} \times \left(\frac{1}{4} - \frac{2}{5} + \frac{3}{2}\right)$

3 $\left(\frac{2}{3}\right)^{2} + 1 - \frac{4}{9} \div \frac{2}{3} - \left(\frac{4}{9} + \frac{2}{3}\right) \times \frac{1}{2}$

Types of fraction



Conversion of fractions and decimals

A fraction can be converted into decimals by dividing its numerator by the denominator and writing the answer as a decimal.

Example: $\frac{6}{5} = 6 \div 5 = 1.2$

Some fractions, called **recurring decimals**, do not divide easily, and they give the same number or series of numbers to infinity.

Example: $\frac{2}{3}$ = 0.6666666........ or 0.6 recurring

In most of such cases, you can give your answer up to two or three or a specified number of decimal places. However, if you are giving an answer to a number of decimal places, the decimals have to be appropriately rounded off.

In general, if the recurring decimal has 5 or a higher number recurring, you need to round up the last digit to the next number

Example: $\frac{2}{3}$ = 0.6666666....... can be rounded off as 0.667 if the answer is expressed up to three decimal places.

If the recurring decimal has less than 5 as the number recurring, you need to leave the number as is.

Example: $\frac{1}{3}$ = 0.33333..... = 0.33 to two decimal places

A decimal can be converted into a fraction by removing the decimal and dividing the number by a multiple of 10 equal to the number of digits after the decimal place.

Example: $1.25 = \frac{\cancel{125}^{5}}{\cancel{100}^{4}} = \frac{5}{4} = 1\frac{1}{4}$

Or

$0.03 = \frac{3}{100}$

OVER TO YOU

Activity 5

Give two examples of each of the following:

1 Equivalent fractions

2 Recurring decimals

OVER TO YOU

Activity 6

Convert the following fractions into decimals.

1 $1\frac{4}{5}$

2 $\frac{7}{11}$

3 $\frac{16}{3}$

4 $3\frac{3}{5}$

5 $\frac{78}{13}$

OVER TO YOU

Activity 7

Solve and express your answers as a fraction and a decimal to two decimal places.

1 $\frac{1}{4} + 2\frac{6}{11} - \frac{2}{5} \div \frac{8}{10} \times \frac{1}{4}$

2 $4\frac{3}{4} + 1.2 - 3.5 - \left(\frac{6}{7} \times 14.7\right) + 3\frac{1}{2}$

1.2 Express numbers in various forms for making comparisons

Expressing numbers in the standard form

- A number can be expressed in terms of its **exponent** or index of its power. Therefore, 3×3×3×3×3 can be expressed as 3^5, where the exponent or the index of the power "5" is indicated as a superscript at the end of the number.

"Verbally, the expression 3^5 is read as "three to the power five" or "three raised to the fifth power"."

- An exponent can also be negative. In such cases, you would express the number as the **reciprocal** of the number to the positive power of the exponent.

  Therefore, $4^{-3} = \frac{1}{4^3} = 0.156$

- When we have very small or very large numbers, you can express these in a more manageable form called **standard form** $A \times 10^n$, where n is an integer and A is a number less than 10.

For example:

87,038,993,020,000 can be expressed in standard form by multiplying it by 10s raised to the required power to make it manageable.

Therefore, 87,038,993,020,000 can be simplified to $87{,}038 \times 10^9$ as

$$87{,}038{,}000{,}000{,}000 = 87038 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10$$

"In standard form, a number × 10^n can be expressed as number being multiplied n times by 10."

The number can be expressed in decimal terms in standard form by adding a decimal point and each time the decimal is moved left the power of 10 is increased by 1.

Therefore, the number can also be expressed as:

870.38×10^{11}

or

87.038×10^{12}

or

8.7038×10^{13}

This final form is the number in standard form as 8.703 is less than 10.

This makes the number more manageable for applying arithmetic operations and also for comparing with other numbers. Of the two numbers, the number that is raised to a higher power is larger.

A number × 10^{-n} (a negative exponent) implies that the number is being multiplied by $\frac{1}{10^n}$ where n is the value of the exponent.

Therefore,

$0.956 = 956 \times 10^{-3} = 956 \times \frac{1}{10^3}$

8703.8×10^n

or

$870.38 \times 10^{n+1}$

or

$87.038 \times 10^{n+2}$

Decimal moves left by one position, power goes up by 1

OVER TO YOU

Activity 8

Express the following in the standard form $A \times 10^n$

1 4562.3 where $n = 5$

2 2,893,000,000 where $n = 7$

3 623.54 where $n = -3$

4 875.23 where $n = -1$

OVER TO YOU

Activity 9

Express the following numbers in full.

1 505.05×10^{-2}

2 8.01×10^{3}

3 635.25×10^{7}

4 8745×10^{-3}

Expressing numbers as a percentage of another number

Per cent is derived from the Latin phrase "per centum" which means "by the hundred" or "per hundred". More commonly it is written as %. Therefore, 30% means "30 per hundred" or "30 out of 100".

You can express quantities in percentages when these need to be compared by bringing them to a common base. **This base is 100**. For example, 30% is smaller than 32%.

You can express per cent as a whole number or convert it into a fraction or a decimal.

- You can convert it into a fraction by dividing it by 100 and removing the % sign. Removing the % sign simply implies dividing the number by 100.

  Example: $30\% = \frac{30}{100} = \frac{3}{10}$

- You can convert it into a decimal by removing the decimal point and moving the decimal point two places to the left.

  Example: $30\% = 0.30$

  You can convert a whole number or a decimal or a fraction into percentage by simply multiplying it by 100 and putting a % sign.

  Example: $2 = 2 \times 100 = 200\%$

  $\frac{1}{8} = \frac{1}{8^{2}} \times \cancel{100}^{25} = \frac{25}{2}\% = 12.5\%$

  $0.3 = 0.3 \times 100 = 30\%$

OVER TO YOU

Activity 10

Express the following as percentages.

1 $\frac{3}{5}$

2 $\frac{8}{15}$

3 $\frac{1}{30}$

4 $3\frac{6}{7}$

OVER TO YOU

Activity 11

Express the following as fractions.

1 34.62%

2 84.1%

3 103.20%

4 55.5%

Rules applying to percentages

1 You can add or subtract percentages that are part of the same whole.

Example 1

For example, 40% of the students in a class study financial management, 25% study human resource management and the remaining students study marketing. Here the whole (100%) is the total number of students in the class. Therefore, you can calculate the total percentage of students studying marketing as:

100% – (40% + 25%)

= 100% – 65%

= 35%

However, you cannot add percentages that lie outside the group of numbers being considered.

2 To calculate the specific percentage of number, simply convert the percentage to be calculated into a fraction and then multiply the number by that fraction.

Example 2

David earns a net amount of $32,000 per year. He saves 15% of this. How much does David save?

Since David saves 15% of his net earnings, we need to calculate 15% of $32,000.

$15\% \times 32{,}000 = \frac{15}{100}$ 15/100 × 32,000 = $4800

David's friend Salim saves 12% of his annual earnings of $35,000. Who saves more money every year – David or Salim?

Salim saves 12% of $35,000.

$12\% \times 35{,}000 = \frac{12}{100} \times 35{,}000 = \$4200$

Therefore, David saves more earnings every year.

3 Percentages can only be calculated for numbers that have the same units of measurement. The units must be made similar before the calculation is done.

Example 3

If you want to express 5000 grams as a percentage of 200 kilograms, then either grams should be converted into kilograms or kilograms as grams.

Since 1 kilogram = 1000 grams

200 kilograms = 200,000 grams

Therefore, 5000 grams as a percentage of 200,000 grams is:

$\frac{5000}{200{,}000} \times 100 = 2\frac{1}{2}\%$ or 2.5%

4 The original value can be calculated given the increase or decrease percentage as follows:

Example 4

Madhu sold 3000 units in 2015. In 2016, units sold increased by 20%. How many units did he sell in 2016?

Number of units in 2016 can be calculated as:

$(100 + 20)\% \times 3000 = \frac{120}{100} \times 3000 = 3600$

5 You can calculate the increase or decrease percentage as follows, given the original and the new values.

Example 5

The cost of manufacturing one unit of product increased from $5 to $8. What was the percentage increase in cost?

The increase in cost = $8 – $5 = $3

$\text{Increase in cost } \% = \frac{\text{Increase in profit}}{\text{Original profit}} \times 100$

$= \frac{3}{5} \times 100 = 60\%$

Example 6

The profit of a business decreased from $15,000 to $12,000. What was the percentage change in profit?

$\text{Decrease in cost } \% = \frac{\text{Decrease in profit}}{\text{Original profit}} \times 100$

$\frac{15{,}000 - 12{,}000}{15{,}000} \times 100 = \frac{3000}{15{,}000} \times 100 = 20\%$

6 The total value or part of the total value can be determined from a percentage as follows:

Example 7

60% of students in Rehman's class play some sport regularly. There are 120 students in his class. How many students do not play any sport?

Total students in the class = 120 = 100%

Students who do not play a sport = 100% – 60% = 40%

Therefore, 40% of total students in Rehman's class do not play any sport.

Number of students in Rehman's class who do not play a sport is:

40% × 120

$= \frac{40}{100} \times 120 = 48$

Example 8

Lisa pays $1500 every month as rent. This is 25% of her monthly income. What is Lisa's total monthly salary?

If 25% of the salary = $1500

100% of the salary = $\$\frac{1500}{25} \times 100 = \$6000$

OVER TO YOU

Activity 12

Solve the following question.

In a class of 150 students, 70% have opted for business statistics, 50% students study economics. Work out how many students study both the subjects.

OVER TO YOU

Activity 13

1 Express 45 minutes as a percentage of an hour.
2 Express 750 grams as a percentage of 700 kilograms.
3 Express 240 minutes as a percentage of an hour.

OVER TO YOU

Activity 14

Solve the following question.

Carole pays a rent of $2000 per month. The landlord decides to increase her rent to $2300 per month from August. Calculate the percentage increase in the rent that Carole will pay from August.

Comparison of numbers using ratios and proportions

Ratios

A **ratio** expresses the relationship between two or more values. The values or quantities can have a ratio only if they have the same units of measurement. The symbol ":" (colon) denotes ratio. Therefore, two numbers x and y , when written as $x:y$ indicate the ratio between x and y.

Let's understand ratio with the example below:

John plays tennis for one hour every day and Susan plays tennis for 90 minutes. What's the ratio between the time spent by John to the time spent by Susan?

The steps to calculate the ratio are:

Make the units of measurement same. Note that the question states one hour for John and 90 minutes for Susan. Therefore, first convert either John's one hour into minutes or Susan's minutes into hours. Let's convert one hour into 60 minutes. Therefore,

$$\frac{60 \text{ minutes}}{90 \text{ minutes}} = \frac{60}{90} = \frac{2}{3}$$

Here, 2 (the numerator) is called an **antecedent** and 3 (the denominator) is called **consequent**.

When expressed as a ratio, the resulting fraction $\frac{2}{3}$ is written as 2:3. This ratio indicates that for every 2 minutes of tennis played by John, Susan has played 3 minutes. In other words, Susan has spent one and a half times more time on playing tennis than John.

A ratio can also be expressed in various ways as shown in Table 6.

Ways to express ratios	Example
Ratio between whole numbers	2:3
Ratio between fractions	$\frac{1}{3}:\frac{2}{3}$
Ratio between decimals	2.3:4.5

Table 6: Ways to express ratios

REVISION on the go

Ratios are used in many business situations. With the use of ratios, various related quantities can be quickly compared. For example, the ratio of sales done by a business in the years 2015 and 2016 can instantly tell us whether there is a positive improvement or decline in sales.

Ratios: general rules

There are some important rules for ratios. These are shown in Table 7.

Rules for ratios	Example
A ratio can only be expressed for items that have some form of comparable measurement.	The ratio between a car and a house is possible only if the cost of both items is being compared (cost is comparable measurement for two dis-similar items).
If the numerator and the denominator of a ratio is multiplied or divided by the same number, there is no change in the ratio.	$\frac{4}{5} \times \frac{2}{2} = \frac{4 \times 2}{5 \times 2} = \frac{8}{10} = \frac{4}{5}$
Ratios should always be reduced to the simplest form.	The ratio between 60 and 90 is 2:3.
Two ratios can be compared by reducing the ratios in such a form that the first term of the ratio is the same.	Compare the two ratios 4:3 and 5:4 $4:3 = \frac{4}{4}:\frac{3}{4} = 1:\frac{3}{4}$ $5:4 = \frac{5}{5}:\frac{4}{5} = 1:\frac{4}{5}$

Table 7: Rules for ratios

REVISION on the go

Dividing a number using the given ratio

A ratio can be used to divide a number or a value. If there are two terms in the ratio dividing a number, there will be two fractions that we will multiply the number with separately, to get the two parts. If there are three terms then there will be three fractions and so on. If a number is to be divided into two parts:

Step 1: Add the two terms of the ratio and use this as the denominator of the two fractions you will use.

Step 2: Take the first term of the ratio as the numerator in the first fraction.

Step 3: Multiply the number to be divided by the first fraction to get the first part of the division.

Step 4: Take the second term of the ratio as the numerator in the second fraction and multiply this second fraction by the number to get the second part.

The same approach can be used in the case of ratio of more than three values or quantities. This is illustrated in the following example.

Example

You enter into a partnership with Khalid and Maureen and decide to share profits and losses in the ratio 3:2:1. The partnership made a profit of $48,000 in the year. What will be the share of each partner in the profit?

Solution: There are three terms in the ratio that should be used for dividing $48,000. Therefore, there will be three fractions, and these will be used to divide the profit.

Denominator of the fractions = 3 + 2 + 1

- The first ratio representing your share will be $\frac{3}{6}$
- The second ratio representing Khalid's share will be $\frac{2}{6}$ and
- The third ratio representing Maureen's share will be $\frac{1}{6}$

Therefore,

Your share = $\frac{3}{6} \times 48{,}000 = \$24{,}000$

Khalid's share = $\frac{2}{6} \times 48{,}000 = \$16{,}000$

Maureen's share = $\frac{1}{6} \times 48{,}000 = \$8000$

OVER TO YOU

Activity 15

Express the following in the simplest form of the ratio.

1. 20 is to 60
2. 55 is to 220
3. 90 is to 75

OVER TO YOU

Activity 16

1. Divide 360 in the ratio of 3 : 2 : 1
2. Divide 4500 in the ratio 6 : 5 : 4
3. If A, B and C share profits in the ratio 2 : 3 : 4, work out B's share if the profit of the business is $36,000.

OVER TO YOU

Activity 17

Chao and Min enter into partnership sharing profits and losses in the ratio 2 : 3. Lynn joins the partnership and the partners decide to share the profit equally.

The partnership makes a profit of $30,000. What will be difference in the shares of Chao's profit before and after Lynn's admission to the partnership?

OVER TO YOU

Activity 18

Aamir had $32,000 in cash. He invested 8.5% in stocks and shares and gave 25% to his parents. How much did he invest in stocks and shares? How much did he give to his parents? What is the amount left with him?

Proportions

Ratios and proportions are inter-related. Two quantities are in direct proportion if they increase or decrease in the same ratio.

A **proportion** is a mathematical statement with two equal ratios. It is written a "::" and is read as "as"

For example: 5 : 10 :: 1 : 2 is an expression of a proportion. Stated in words, it reads:

"5 is to 10 as 1 is to 2"

The fractions that this expression represents are:

$$\frac{5}{10} = \frac{1}{2}$$

In a proportion, 5 : 10 :: 1 : 2 the first and the last terms, that is, 5 and 2 are called "**extremes**" and the two middle terms, that is, 10, and 1 are called "**means**".

In a proportion, the product of the extremes = the product of the means. This is a very important principle that is used for calculating the missing values in a given expression of proportion.

Example: Udoka spends $6 to buy 12 pens. How much will 18 pens cost?

This problem can be expressed as: $6:12::?:18$

The value of "?" can be determined as follows:

$6:12::?:18$

$$\frac{6}{12} = \frac{?}{18}$$

? × 12 = 6 × 18

$$? = \frac{6 \times 18}{12} = 9$$

Therefore, 18 pens will cost $9.

OVER TO YOU

Activity 19

Find the missing values.

1 2 : 3 :: 5 : ?

2 3.5 : 7.0 :: ? : 5.4

3 6 : ? :: 12 : 18

4 ? : 4 :: 4 : 8

PROPORTIONS: GENERAL RULES

There are some general rules that apply to proportions where terms of proportions are: *r*, *s*, *t* and *u*.

1 If $\frac{r}{s} = \frac{t}{u}$ then the inverse of these, that is, $\frac{s}{r} = \frac{u}{t}$ is also true

2 If $\frac{r}{s} = \frac{t}{u}$ then $\frac{r}{t} = \frac{s}{u}$

3 If $\frac{r}{s} = \frac{t}{u}$ then $\frac{r+s}{s} = \frac{t+u}{u}$

4 If $\frac{r}{s} = \frac{t}{u}$ then $\frac{r-s}{s} = \frac{t-u}{u}$

5 If $\frac{r}{s} = \frac{t}{u}$ then $\frac{r+s}{r-s} = \frac{t+u}{r-s}$

OVER TO YOU

Activity 20

Find the missing values.

1 $\frac{0.60}{0.75} = \frac{0.48}{x}$

2 $\frac{5}{2} = \frac{x}{1200}$

3 $\frac{16}{x} = \frac{x}{49}$

1.3 Performing simple financial calculations

This section is focused on mathematical concepts that are specific to business situations. Four independent topics that are covered in this section include:

1 Simple and compound interest

2 Discounted or present value of money

3 Depreciation of an asset

4 Miscellaneous business calculations such as calculations of wages and taxation, discounts, foreign exchange conversions

We will understand these topics by performing calculations for given business situations.

Calculation of simple and compound interest

Interest relates to money borrowed or invested. It is the amount in addition to the original amount that the lender receives from the borrower while the original amount is being used or returned.

There are four key terms for understanding interest:

- **Lender:** A person or institution (such as a bank) that offers the loan.
- **Borrower:** A person or institution that receives the loan.
- **Principal:** The amount of loan.
- **Interest:** The amount that the borrower has to pay to the lender in addition to the repayment of the principal.

Therefore,

repayment of loan = principal + interest

Note: An ordinary person who deposits their money in a savings bank account or a fixed duration account (i.e. invests money), becomes the lender to the bank, and receives (earns) interest while the bank uses their money for productive purposes. The reverse is also true. The person who borrows money (i.e. takes a bank loan) from the bank pays interest to the bank.

The interest has certain characteristics:

- It can be either **simple interest** or **compound interest**.
- It is mostly calculated as a pre-determined percentage of the principal.
- When the rate of interest is given as a percentage per year it is called the annual interest rate or interest rate per annum.

Simple interest

Simple interest is the interest that applies only to the amount borrowed or invested. It is calculated as a percentage of the principal amount only for a number of years. It stays the same for the duration of the loan (or investment). Normally, short duration loans charge a simple interest and short-term investments earn a simple interest. An interest of 10% per annum on $100 means that $10 is charged as interest on $100 at the end of the year. Therefore, the total amount due to the lender at the end of one year will be $110.

The duration of the loan (or investment) is the time that the principal amount is either borrowed or invested. The duration is generally given in years but can also be stated in months or days.

Simple interest can be calculated with the formula:

$$S.I. = \frac{P \times R \times T}{100}$$

Where,

$S.I.$ = Simple interest

P = Principal amount

R = Rate of interest per annum

T = duration of the loan or investment

CASE STUDY 1

Monica's loan

Monica wants to renovate her office. She wants to finance the renovation partly with her own money and the remaining by borrowing a reasonable amount from her friend.

She is considering taking a loan of $8000 from her friend for 5 years at a simple interest of 12% per annum. What amount will Monica pay back to her friend at the end of 5 years if she takes the loan?

In Case study 1: Principal = $8000

Duration = 5 years

Rate of interest = 12%

To calculate the amount of money that Monica will have to repay, we should calculate the simple interest using the formula:

$$S.I. = \frac{P \times R \times T}{100}$$

This is a straightforward calculation:

$$S.I. = \frac{8000 \times 12 \times 5}{100}$$

$$= \$4800$$

Since $4800 is the simple interest, Monica needs to pay back $8000 + $4800 = $12,800 to her friend at the end of 5 years.

CASE STUDY 1 (CONTINUED)

Monica thinks more and decides that she does not want to pay more than $3000 interest at 12% per annum to her friend at the end of 5 years. She is thinking of reducing the amount so that the interest does not exceed $3000.

What amount should she borrow so that her interest is not more than $3000?

This time Monica knows that $S.I.$ = $3000, R = 12% and T = 5 years. She needs to find P.

We will adapt the formula $S.I. = \frac{P \times R \times T}{100}$ and state it in terms of P to find our answer.

The terms of this formula can be transposed as:

$$\frac{100 \times S.I.}{R \times T} = P$$

or

$$P = \frac{100 \times S.I.}{R \times T}$$

Substituting the values:

$$P = \frac{100 \times 3000}{12 \times 5}$$

= \$5000

Therefore, Monica should not borrow more than \$5000 if she does not want the interest to exceed \$3000.

OVER TO YOU

Activity 21

Salman takes a loan for \$4500 for 3 years at a rate of 7% rate of simple interest. Work out the amount of interest Salman will need to pay after three years.

CASE STUDY 1 (CONTINUED)

Monica's loan

Monica finds out that the bank can lend her \$8000 at an interest rate of just 8% simple interest per annum for a longer duration. Monica can pay interest up to \$3200. What duration of loan should she decide so that the interest on the loan is \$3200?

In this instance, $S.I.$ = \$3000, R = 8% and P = \$8000. Monica would like to know the value of T.

Just like earlier, the formula terms of the formula $S.I. = \frac{P \times R \times T}{100}$ are transposed as:

$$\frac{100 \times S.I.}{R \times P} = T$$

or

$$T = \frac{100 \times S.I.}{R \times P}$$

Substituting the values:

$$T = \frac{100 \times 3200}{8 \times 8000}$$

= 5 *years*

Therefore, the best duration of loan for Monica is 5 years. Using this, she can avoid paying interest of more than $3200.

Simple interest formula

$$S.I. = \frac{P \times R \times T}{100}$$

REVISION on the go

OVER TO YOU

Activity 22

Calculate the rate of simple interest Monica should pay if she borrows $8000 for 5 years and only $2800 interest.

OVER TO YOU

Activity 23

Noel invests $5000 in a high interest paying bank account. The bank will pay a simple interest at 10% per annum on a six-month basis into this account. After how many years will the total amount in the account double?

Compound interest

Unlike simple interest, compound interest is the interest calculated not only on the principal amount but also on the accumulated interest. At the end of year 1 there is no difference between the simple interest and compound interest as there is no accumulated interest. However, in year 2, the principal for compound interest will be equal to the sum of the principal and the interest accumulated for year 1. Therefore, interest will be calculated on a higher base. This happens with every passing year till the end of the term.

Let's understand compound interest with an example:

Example: Jiao takes a loan of $1000 for 3 years on 10% compound interest.

At the end of year 1, the interest is calculated as:

$$\frac{\$1000 \times 10 \times 1}{100} = \$100$$

At the end of year 2, interest will be calculated on \$1100, i.e. principal (\$1000) and the accumulated interest for year 1 (\$110). Therefore, the interest will be:

$$\frac{\$1100 \times 10 \times 1}{100} = \$110$$

At the end of year 3, interest will be calculated on \$1210, i.e. principal (\$1000) and the accumulated interest for year 1 and 2 (\$210).

These year-on-year calculations for compound interest can be done with the formula:

$$C.I. = P\left(1 + \frac{R}{100}\right)^T - P$$

Where,

$C.I.$ = Amount

P = Principal amount

R = Rate of interest per annum

T = Duration of the loan or investment

The amount (A) accrued at the end of the period is $P\left(1 + \frac{R}{100}\right)^T$

Compound interest formula

$$C.I. = P\left(1 + \frac{R}{100}\right)^T - P$$

REVISION on the go

Illustration of compound interest for 4-year period.

End of Year 1: Interest applied only on principal, therefore, compound interest = simple interest

↓

End of Year 2: Interest applied on principal plus interest accumulated for year 1

↓

End of Year 3: Interest applied on principal plus interest accumulated for year 1 and 2

↓

End of Year 4: Interest applied on principal plus interest accumulated for year 1, 2 and 3

REVISION on the go

Let's understand how to make a decision on borrowing with simple and compound interest with the following case study.

CASE STUDY 1 (CONTINUED)

Monica's loan

Monica's bank offered to lend her $8000 for a 5-year period at interest of 7% compounded annually.

Monica does not know whether this is a better offer than the earlier 8% simple interest.

Monica should compare the total amount repayable after 5 years at 7% compound interest against the amount payable at 8% per annum simple interest.

We know that the amount payable after 5 years with simple interest is:

$$\$8000 + \$3200 = \$11{,}200$$

Now let's calculate the amount payable when interest is compounded annually with the formula:

$$A = P\left(1 + \frac{R}{100}\right)^T$$

We know:

P = $8000

R = 7%

T = 5 years

Substituting the values:

$$A = 8000\left(1 + \frac{7}{100}\right)^5$$

$$A = 8000\left(\frac{100 + 7}{100}\right)^5$$

$$A = 8000\left(\frac{107}{100}\right)^5$$

$$= \$11{,}220.41$$

We know that the amount payable after 5 years with compound interest is: $11,220.41

Note that even though the rate of interest is 1% lower, the total amount payable at the end of 5 years with compound interest (7%) is $20.41 more than the amount payable with simple interest (8%).

CASE STUDY 2

Akilah deposits $5000 in a bank account for two and a half years. The bank pays an interest at 10% per annum. The interest is calculated after every 6 months and is automatically deposited into Akilah's account.

Akilah wants to calculate the total amount that will be there in her bank account at the end of the period. She also wants to know the interest component.

Case study 2 has two aspects:

1 The bank calculates and pays interest every six months. The rate of interest is 10%, therefore every six months the bank pays 5% interest (half of 10%) into the bank account.

   Therefore, $R = 5\%$

2 The total period of the deposit is two and a half years. This implies that Akilah's bank account will receive the interest 5 times.

   Therefore $T = 5$

Use the formula $P\left(1 + \frac{R}{100}\right)^T$ to determine the amount due on maturity.

$5000\left(1 + \frac{5}{100}\right)^5$

$= 5000 \times \left(\frac{105}{100}\right)^5$

$= \$6381.41$

Therefore, Akilah should have $6381.41 in her bank account at the end of two and a half years.

Of this total amount, the interest earned by her amounts to $1381.41.

OVER TO YOU

Activity 24

Which is greater?

a A sum of money accumulating 3% compound interest for 12 years; or

b The same sum of money accumulating 6% compound interest for 6 years.

Calculation of discounted or present value of money

The discounted or **present value of money** is closely related to the concept of interest. The concept of present value is used to calculate the worth of future sums of money in the present day based on a specified discount rate. The higher the discount rate, the lower is the present value of future cash flows. Instead of determining the future value of an investment made today at a specific rate of interest, it determines the present value of the future value of an investment at the required rate of return.

The concept of present value is one of the fundamental concepts in finance. It is widely used in financial modelling for taking long-term investment decisions. Present value accounts for the time value of money. In other words, it accounts for the return that one can get on the money invested today.

Present value can be calculated as follows:

$$PV = \frac{CF}{(1 + R\%)^T}$$

Where,

PV = Present value

CF = Cash flow

R = Required rate of return or interest or discount rate

T = Number of periods

CASE STUDY 3

Brian is evaluating investing in a project today. The project will result in a one-time cash inflow of $90,000 at the end of three years. He estimates the discount rate to be 12% and wants you to help him find the present value of this cash flow.

Brian can calculate the present value of $90,000 as follows:

$$PV = \frac{90{,}000}{(1 + 0.12)^3}$$

$$= \frac{90{,}000}{1.12^3} = \$64{,}060.22$$

It is NOT necessary that investment is always made into a project with the aim of getting consolidated cash inflow at the end of the project. Often investments are made for getting periodic cash inflows, typically on an annual basis. In many cases, a project may give the investor the same amount of cash inflow every year. The present value in such cases can be determined by using the following formula:

$$PV_A = CF_A \times \frac{\left(1 - \left(\frac{1}{1 + R\%}\right)^T\right)}{R\%}$$

Where:

PV_A = the present value with annual cash flow

CF_A = Cash flow each year

R = Required rate of return or interest or discount rate

T = Number of periods

CASE STUDY 3 (CONTINUED)

Brian also has a second option that will give him $30,000 cash inflow every year for the next three years. He wants to understand whether this option is better than receiving $90,000 as a one-time cash inflow at the end of the project period.

In this instance, the present value can be calculated as:

$$PV_A = 30{,}000 \times \frac{\left(1 - \left(\frac{1}{1 + 0.12}\right)^3\right)}{0.12}$$

$$= 30{,}000 \times \frac{\left(1 - \left(\frac{1}{1.12}\right)^3\right)}{0.12}$$

$$= 30{,}000 \times \frac{(1 - 0.712)}{0.12}$$

$$= 30{,}000 \times \frac{0.288}{0.12}$$

$$= \$72{,}000$$

It is evident that the present value of the project with annual returns is higher than the one-time cash inflow at the end of three years. Therefore, Brian should decide on getting annual returns.

$$PV = \frac{CF}{(1 + R\%)^T}$$

REVISION on the go

OVER TO YOU

Activity 25

Calculate the present value of receiving $1500 every year if you receive the amount for the next:

a 6 years and the discount rate is 12%

b 5 years and discount rate is 10%

c 6 years and the discount rate is 10%

d 5 years and the discount rate is 12%

Calculation of depreciation of an asset

A business has many assets that are used in operations. Examples include machinery, computers, furniture, cars and vans among others. Over time, these assets suffer from wear and tear; as a result, their monetary value declines. This is called depreciation.

Depreciation is defined as the reduction in the value of an asset over time due to wear and tear, obsolescence (becoming out of date), technological changes and several other factors. It is an important concept for any business because the principles of accounting demand that the assets should be assessed at their real worth, i.e. their value after calculating depreciation.

The amount of depreciation charged on an asset depends on:

- The useful life of the asset or the estimated period for which the asset will be used;
- Salvage or scrap or residual value of the asset or the value at which the asset can be sold at the end of its useful life;
- The method of depreciation used.

There are various methods of calculating depreciation. Two commonly used methods of depreciation are:

- Straight-line method;
- Reducing balance method.

This section will discuss these two methods.

Straight-line method

Straight-line method of depreciation spreads the cost of the asset evenly throughout the useful life of the asset. It assumes that there will be a uniform use of the asset over its life. It is calculated as:

$$\text{Annual depreciaton} = \frac{\text{Cost of asset} - \text{Scrap or salvage value}}{\text{Useful life}}$$

The absolute depreciation charged each year stays the same.

Example 1

Ajit purchases a machine on 1 January 2015 for $11,000. The machine has a useful life of 5 years. At the end of its useful life, the machine can be sold off for $1000. Depreciation on the machine by applying the straight-line method is calculated as:

$$\text{Annual depreciaton} = \frac{(\$11,000 - \$1000)}{5} = \frac{\$10,000}{5} = \$2000$$

Therefore, the value of the machine will be:

- At the end of year 1 : $11,000 – $2000 = $9000
- At the end of year 2 : $9000 – $2000 = $7000
- At the end of year 3 : $7000 – $2000 = $5000
- At the end of year 4 : $5000 – $2000 = $3000
- At the end of year 5 : $3000 – $2000 = $1000

Example 2

Emma purchased office equipment at a cost of $5400. This office equipment will have no residual value at the end of its life. She charges depreciation of $540 every year. After how many years will her office equipment have no value?

Note that since the office equipment will have zero value when it is completely depreciated.

Number of years within which it will be completely depreciated = $\frac{5400}{540}$ = 10 years.

Example 3

What is the rate of depreciation charged by Ajit and Emma?

$$\text{Rate of depreciation} = \frac{\text{Depreciation per annum}}{(\text{Cost} - \text{Scrap or salvage value})} \times 100$$

$$\text{Therefore, in Ajit's case, the Rate of depreciation} = \frac{2000}{(11{,}000 - 1000)} \times 100 = 20\%$$

$$\text{In Emma's case Rate of depreciation} = \frac{540}{(5400 - 0)} \times 100 = 10\%$$

OVER TO YOU

Activity 26

A factory purchases a machine for $40,000 in January. The machinery is estimated to have a useful life of 5 years after which it will be sold for $2000. Calculate the rate of depreciation the business should charge every year.

Reducing balance method

This method is also known as **diminishing balance method**. Under this method, depreciation for each year is calculated at a fixed rate on the diminishing balance of the value of the asset.

> *In reducing balance method, the final balance of the asset is never zero.*

The method charges a higher depreciation in the earlier years of the useful life of the asset.

Annual depreciation = (Net Book Value – Scrap or salvage value) × Rate%

Net book value is the value of the asset at the start of the year. It is calculated by deducting the total depreciation charged on the asset till that point from the cost of the asset.

Example 4

Aadil purchases a motor vehicle on 1 January 2015 for $12,000. He charges a depreciation of 15% using the reducing balance method. What will be the value of the motor vehicle on 1 January 2018?

2015

Depreciation in 2015 = 12,000 × 15% = $1800

2016

Net book value on 1 January 2016 = $12,000 – $1800 = $10,200

Depreciation in 2016 = 10,200 ×15% = $1530

2017

Net book value on 1 January 2017 = $10,200 – $1530 = $8670

Depreciation in 2017 = 8670 × 15% = $1300.50

2018

Value on 1 January 2018 = $8670 – $1300.5 = $7369.50

Depreciation can easily be calculated with the formula.

$DV = C(1–R)T$

Where,

DV = depreciated value at the end of the period T

C = Original cost of the asset

R = Rate of depreciation

T = Time duration or period

In Example 4, value at the beginning of 2018 is the same as value at the end of 2017 (31 December 2017). As the asset was purchased on 1 January 2015, the depreciation for three years would have been charged on 31 December 2017. Therefore, depreciated value on 31 December 2017 will be:

$DV = 12{,}000(1 - 0.15)^3$

= $7369.50

Formula: Straight-line depreciation

$$\text{Annual depreciaton} = \frac{\text{(Cost of asset – Scrap or salvage value)}}{\text{Useful life}}$$

Formula: Reducing balance depreciation

Annual depreciaton = (Net Book Value – Scrap or salvage value) × Rate%

REVISION on the go

OVER TO YOU

Activity 27

Malcolm depreciates his computer at a rate of 35% each year. At the end of 2016 the value of the computer was $5740. Work out the value of the computer at the end of 2014 based on reducing balance method.

OVER TO YOU

Activity 28

Jesse bought a new car. She depreciates the car at 10% using the reducing balance method.

After how many years will the value of her car be less than 50% of the original value?

Other business calculations

Every business must do calculations for wages and salaries, and discounts to customers. Some import-export driven businesses have to calculate foreign currency conversions on a regular basis. This topic discusses the numerical calculations involved in such business activities. The technical or legal aspects of these calculations are beyond the scope of this syllabus.

Wages and salaries (including deductions for taxation, if any)

Employees receive wages and salaries. Some employees get fixed remuneration every month. We can calculate annual salaries by multiplying the monthly salary by 12.

There are other employees who work under different contractual agreements such as:

- receive a fixed monthly salary + commission;
- receive wages based on the number of hours worked;
- receive wages based on units produced, etc.

In many countries, deductions are made from the salaries and wages for pension contributions or taxation. In many organisations, employees who work more than the required number of hours are paid overtime wages.

Overtime wage rate is generally higher than the normal wage rate and may be indicated as time and a half or time and a quarter, etc. The various terms involved in these calculations are:

- Time and a half: this means normal hourly rate + 0.5 of normal hourly rate = hourly rate × 1.5
- Time and a quarter: this means normal hourly rate + 0.25 of normal hourly rate = hourly rate × 1.25

- Gross wages or gross salary: this is the total amount due to an employee before any deductions.
- Net wages or net salary: this is the total amount due to the employee after deductions.

Example 1

For the month of October, Jian was paid for 136 hours at $8 per hour and 28 hours at time and a half. The employer deducted $275 towards taxation and $60 for pension contributions from the gross amount. What did Jian's receive in the month of October?

There are two components in Jian's gross remuneration:

- Amount based on number of normal hours worked;
- Amount based on overtime worked.

Therefore, Jian's gross remuneration

$= (136 \times \$8) + 28 \times (\$8 \times 15)$

$= \$1088 + \$336 = \$1424$

Out of this, deduction towards taxation and pension is subtracted to get the net amount received by Jian in October.

$\$1424 – \$275 – \$60 = 1089$

Example 2

Eve is in the sales team in her organisation. She receives a salary of $24,000 per annum. She also receives a commission of 10% on the amount of sales exceeding $7500. In August, she had sales of $10,000. There were no deductions from her salary. What was her salary in August?

There are two components in Eve's remuneration in August

- Fixed amount of monthly salary $\dfrac{\text{Annual salary}}{12} = \dfrac{\$24{,}000}{12} = \$2000$
- Commission on sales over $7500

  $= (10{,}000 – 7500) \times 10\% = \$2500 \times 10\% = \$250$

  Amount received by Eve in August = $2000 + $250 = $2250

OVER TO YOU

Activity 29

Jonathan is paid $5 per unit he manufactures in addition of $1000 every month. In December Jonathan manufactured 150 units. The employer deducted $150 towards taxation and $40 towards pension contributions from the gross amount. Work out the amount Jonathan received in December.

Discounts

Discounts are common in businesses all over the world. A discount is the amount of money deducted from the cost or the face value or the selling price of an item. A business may offer trade discount to attract more customers or encourage them to buy larger quantities. A business may also offer cash discount to encourage its credit customers to make prompt payments.

All discounts can be simply calculated using the following formula:

Discount = List price × Rate%

Example

Anwar bought office furniture that carried the list price of $4800. He got a discount of 8%. What is the amount Anwar paid for the furniture?

Amount paid by anwar = List price – Discount

= $4800 – ($4800 × 8%) = $4800 – $384 = $4416

OVER TO YOU

Activity 30

Atul gives a cash discount of 3% to his credit customers who pay within 15 days of purchase. Shahana bought goods for $1500 from Atul on 3 April and settled her account on 10 April. Calculate how much discount she received.

Foreign exchange conversions

Very often, businesses involved in international trade make or receive payments in currencies other than their domestic currency. Many business people travel outside the country on business trips, and require foreign currency for their expenditure. Many businesses across the world prefer United States Dollars (US$) as the currency for transactions. Business travellers need to carry the valid currency based on the country they are travelling to. For these purposes, businesses have to request their banks to convert their domestic money into US$ for the purposes of making payments. In other words, they have to buy US$ to make payments. This is called **foreign exchange**. Broadly, the same applies to receiving payments.

Currencies of various countries keep varying in value against US$ on a daily basis due to the way the buying and selling of currencies takes place in the currency market. For a business doing a transaction worth millions of dollars, even a small fluctuation of one cent in the buying or selling rate can result in a payment or receipt of money that is much more than their estimated amounts. This makes the foreign exchange conversions a very important task in a business.

Let's understand these calculations with the help of examples.

Example 1

Zumba is travelling to Malawi from the United Kingdom. He wants to convert £500 (500 British pound Sterling) into Malawi Kwacha. How many Malawi Kwacha can he get? The market rates are:

£1 = $1.239, and

$1= 725.23 MWK

Zumba needs the relationship between £ and MWK to find the amount he can get. In this example, the relationship between £ and MWK is indirect; US$ is the common factor between £ and MWK. Therefore, the calculation has two parts:

Calculate how many pounds (£) he can receive for $1.

£1 = $1.239

$1.239 = £1

$$\frac{\$1.239}{1.239} = \frac{£1}{1.239}$$

$1 = £0.807

With this derivation, the relationship between the three currencies can be easily expressed as:

$1 = £0.807 = 725.23 MWK

£0.807 = 725.23 MWK

$$£1 = \frac{725.23}{0.807}\text{MWK}$$

Therefore,

$$£1 \times £500 = \frac{725.23}{0.807} \times 500\,\text{MWK}$$

£500 = 44,9337.05 MWK

Example 2

Meena is travelling to Japan. She exchanges US$1500 and receives 169,080 Japanese Yen. On arrival at the destination airport, she converts $200 more and gets 22,300 yen. How many more Japanese yen did she receive per $1 in the first transaction?

Exchange rate in the first transaction

$1500 = 16,9080 yen

$$\$1 = \frac{16{,}9080}{1500}\text{yen}$$

$1 = 112.72 yen

Exchange rate in the first transaction

$200 = 22,300 yen

$$\$1 = \frac{22{,}300}{200}\text{yen}$$

$1 = 111.5 yen

Therefore, the extra yen per $1 in the first transaction is 112.72 – 111.5 = 1.22 yen

OVER TO YOU

Activity 31

If 1$ = 0.931 euro, calculate the value of 100 euro in $.

NEED TO KNOW

There are various types of numbers: natural numbers, whole numbers, integers, fractions and decimals.

According to the rules of numeracy, the operations division and multiplication are performed only after opening brackets and simplifying exponents. However, both division and multiplication come before addition and subtraction.

Arithmetic operations, when performed on two positive numbers always result in a positive number.

When arithmetic operations involve either a positive number and a negative number or two negative numbers, the result is based on certain rules.

There are various types of fractions: equivalent fractions, improper fractions, mixed fractions.

An improper fraction can be changed into a mixed fraction by dividing the numerator with the denominator.

A fraction should always be simplified and should always be expressed as the smallest value of the numerator and the denominator.

A fraction can be converted into decimals by dividing its numerator by the denominator and writing the answer as a decimal.

The number can be expressed in decimal terms in standard form by adding a decimal point and each time the decimal is moved towards left the power of 10 is increased by 1.

Quantities are expressed in percentages when these need to be compared by bringing them to a common base.

Ratios express the relationship between two or more values. A ratio can only be expressed for items that have some form of comparable measurement.

Ratios and proportions are inter-related. Two quantities are in direct proportion if they increase or decrease in the same ratio.

Simple interest is the interest that applies only to the amount borrowed or invested. It is calculated as a percentage of the principal amount only, for a number of years.

Compound interest is the interest calculated not only on the principal amount but also on the accumulated interest.

The concept of present value accounts for the time value of money and is widely used in financial modelling for taking long-term investment decisions.

Two commonly used methods of depreciation are straight-line method and reducing balance method.

Straight-line method of depreciation spreads the cost of the asset evenly throughout the useful life of the asset. It assumes that there will be a uniform use of the asset over its life.

Reducing balance method of depreciation for each year is calculated at a fixed rate on the diminishing balance of the value of the asset.

READING LIST

ABE (2017). Introduction to Quantitative Methods. ABE. ISBN 978-1-911550-13-6

Curwin, J. and Slater, R. (2010). Quantitative Methods: A Short Course. Cengage Learning; 001 (Revised Printing) edition. ISBN 978-1844809059

Pickerden, M. and Wright, C. (2016). Essential Maths Skills for AS/A Level Business. Philip Allan, 3rd Ed. ISBN 978-1471863479

RESOURCES

BBC Bitesize (2015). Maths. http://www.bbc.co.uk/education/subjects/z38pycw

Chapter 2

Algebraic Methods

Introduction

Algebra, like arithmetic, is a branch of Mathematics. While arithmetic uses only numbers to express mathematical relationships, algebra uses letters, numbers, exponents, symbols and arithmetical operators of addition, subtraction, multiplication and division to depict mathematical relationships.

Business managers use algebraic methods for solving a variety of problems. In many situations where unknown elements are present, the application of algebraic methods becomes useful. For example, when a firm introduces a new product, price is an unknown element. The solution to the problem of what price should be set for the product is possible with algebra.

In this chapter, you will learn to use algebraic methods. You will understand how to simplify algebraic expressions and solve simple linear equations, formulate and solve simultaneous equations and also quadratic equations. You will also learn to derive the equation of a straight line to show relationship between variables.

Learning outcome

After completing this chapter you will be able to:

2 Apply algebraic methods to formulate and solve business problems. (Weighting 25%)

Assessment criteria

2.1 Simplify or solve equations by employing algebraic methods

2.2 Derive the equation of a straight line to show relationship between variables

2.1 Simplifying or solving algebraic equations

Evaluating algebraic terms and expressions

An algebraic term may be expressed as simply a letter or a combination of a number and a letter, for example, x, xy, $3x$ or $2x^2$. The fundamental rule is that where a term is expressed as a combination of a number and a letter, the number appears before the letter and not vice versa.

An algebraic expression is a collection of algebraic terms used to express a relationship. It uses integers, **constants**, **variables** and arithmetic operators to express a relationship. An example of algebraic expression is:

$$2x^2 + y + 6$$

CASE STUDY 1

Gemma's monthly earnings and expenditure

Gemma works in a small shop on a part-time basis. She earns $100 per day. Every month she spends part of her salary to meet her monthly expenses and pays $20 towards a magazine subscription.

Algebraically, Gemma's monthly earnings and expenditure can be represented as:

$100x - y - 20$

Where,

x = number of days Gemma works in a month

y = monthly expenses

The various elements of this algebraic expression are described in Figure 1.

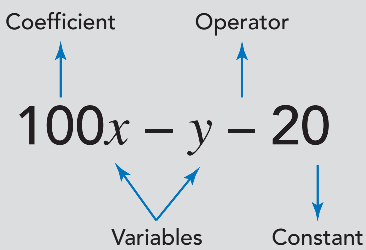


Figure 1: Elements of an algebraic expression

In the algebraic expression in Figure 1, the letters 'x' and 'y' are unknown quantities termed as variables. The values of variables change with the values taken by the unknowns. Number 20 is called a constant because it is a value that is known and does not change.

The **coefficient** of a variable is the number that accompanies a variable. If there is no number placed before the variable, then the coefficient is 1.

OVER TO YOU

Activity 1: Algebraic expressions

Gemma's friend, Kevin, is a web designer. He earns $70 per hour and spends $25 per day travelling to his place of work.

1 Express Kevin's earnings per day as an algebraic expression.

2 Identify the variable, constant and coefficient in the algebraic expression.

Application of arithmetic operations to algebraic terms: general rules

Operations of addition, subtraction, multiplication and division apply to algebraic terms in the same way as they are applied in arithmetic. However, there are some general rules that apply specifically to algebraic terms.

Addition and subtraction

1 Only **like terms** in an algebraic expression can be added or subtracted.

Like terms are the terms that use the same letter and are expressed to the same power, for example, $3x$ and $5x$ are like terms as both use the same letter x, and x is expressed to the same power. Like terms can be added together and/ or subtracted from each other.

Example $x + x - x = 2x$

Example $3x + 5x = 8x$

Based on the same rationale, $3x$ and $3y$ are NOT like terms. Such terms cannot be added together or subtracted from each other.

Example $3x + 3y \neq 6xy$

Some terms are a combination of letters. If the letters in a term do not match the letters in another term then the two terms are NOT like terms, for example, xy and xz are NOT like terms. Similarly, xy and xyz are NOT like terms because the term xyz has an additional variable z. However, xy and yx are like terms.

An algebraic expression that contains multiple terms can be simplified by bringing all like terms together.

Example $x + 2y + 3z - 3y - z + 4x + 4y - 2x$

$= (x + 4x - 2x) + (2y - 3y + 4y) + (3z - z)$

$= 3x + 3y + 2z$

OVER TO YOU

Activity 2: Addition of algebraic terms

Identify the like terms in **each** of the following expressions:

1 $5a^5b^2c + 5ab^2c - 2b^2a^5c + 5abc$

2 $3x^3y + 5xy + 3xy + 3yx^3$

Multiplication

1 Like terms can be multiplied together by raising the product to the power of the number of times the term appears.

Example $x \times x \times x = x^3$

The number 3 in x^3 is called an '**exponent**' or '**power**' and x is called the **base**.

x^1 is equal to x as its exponent is 1. The same is not true for x^0 because the value of $x^0 = 1$

2 Two expressions with the same base can be multiplied even if they have different exponents or powers.

Example $x^2 \times x^3 = (x \times x) \times (x \times x \times x) = x^{(2+3)} = x^5$

Therefore, $x^a \times x^b = x^{(a+b)}$

3 A term raised to the power of another term is denoted with the base raised to the power of product of the two exponents.

Example $(x^3)^2 = x^{(3 \times 2)} = x^6$

or

$(x^2)^{\frac{1}{2}} = x^{(2 \times \frac{1}{2})} = x$

Therefore, $(x^a)^b = x^{ab}$

4 Unlike addition and subtraction, terms that are NOT like terms can also be multiplied and the product is obtained by simply writing the terms being multiplied next to each other.

Example $x \times y \times z = xyz$

5 If one or more terms have coefficients greater than 1, then coefficients are multiplied and the resulting value is written before the product of the letters.

Example $3x \times 2y \times 5z = 30xyz$

General rules in algebra

$x^0 = 1$

$x^1 = x$

$x^2 = x \times x$

$x^3 = x \times x \times x$

$x^a \times x^b = x^{(a+b)}$

$(x^a)^b + x^{ab}$

$x \times y \times z = xyz$

Division

1 Division of two terms can be expressed by using the operator ÷

Example

$4xyz \div zx$

This can also be expressed as:

$$\frac{4xyz}{zx}$$

2 Algebraic terms can be easily cancelled out while dividing. Therefore, if there are like terms in the numerator and the denominator, these can cancel each other out to simplify the algebraic expression.

Example

$$\frac{4\cancel{x}y\cancel{z}}{\cancel{z}\cancel{x}} = \frac{4y}{1} = 4y$$

3 Two expressions with the same base but different exponents can be divided by each other by simply taking the difference between the exponents.

Example

$$\frac{y^5}{y^2} = y^{(5-2)} = y^3$$

Similarly,

$$\frac{4x^5y^2z^4}{z^3x^2y^5} = \frac{4x^{(5-2)}y^{(2-5)}z^{(4-3)}}{z^3x^2y} = 4x^3y^{-3}z = \frac{4x^3z}{y^3}$$

4 A term having a negative exponent can be expressed as its reciprocal with a positive exponent.

Example

$$y^{-3} = \frac{1}{y^3}$$

Example

$$\frac{x^{-\frac{1}{2}}y}{xy^{\frac{2}{3}}} = x^{(-\frac{1}{2}-1)}y^{(\frac{1-2}{3})} = x^{-\frac{3}{2}}y^{\frac{1}{3}} = \frac{y^{\frac{1}{3}}}{y^{\frac{3}{2}}}$$

More general rules in algebra

$$x^{\frac{1}{b}} = \sqrt[b]{x}$$

$$x^{\frac{a}{b}} = \sqrt[b]{x^a}$$

$$\frac{x^a}{x^b} = x^{(a-b)}$$

$$x^{-a} = \frac{1}{x^a}$$

REVISION on the go

There are two more rules that may apply to all algebraic expressions.

$$x^{\frac{1}{b}} = \sqrt[b]{x}$$

$$x^{\frac{a}{b}} = \sqrt[b]{x^a}$$

Thus, $\frac{y^{\frac{1}{3}}}{x^{\frac{3}{2}}}$ can be expressed as:

$$\frac{y^{\frac{1}{3}}}{x^{\frac{3}{2}}} = \frac{\sqrt[3]{y}}{\sqrt{x^3}}$$

OVER TO YOU

Activity 3: Multiplication and division in algebra

Simplify the following expressions:

1 $6a^0$

2 $8x^2yz \times 9z \times y^2$

3 $3t^2 \times 15t^4 \times t$

4 $(4u^2)^3 + 4u^2 \times 4u^3$

5 $x^{-\frac{1}{2}} + x^2y^5z^3 \div xyz^2$

Using brackets in algebra

It is a good practice to simplify all terms within the brackets as much as possible. This makes the algebraic expression more manageable.

There are primarily three key rules that apply to brackets in algebra.

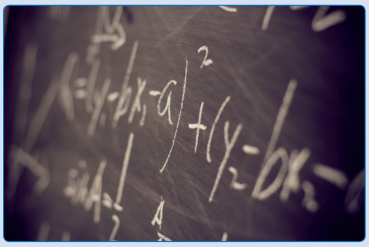
1 Each term within the bracket should be multiplied by the term in front of the bracket.

2 If the bracket is preceded by a + sign, each term within the bracket stays the same.

3 If the bracket is preceded by a – sign, the sign of each term within the bracket changes. Thus a + sign becomes a – sign and a – sign becomes a + sign.

CASE STUDY 2

Using brackets



Fatima wants to simplify the following algebraic expression.

$$2(3x^2 + 5z^3 + \frac{3z^4}{z} + 2zx^3) - (2y + 5x^2 - y)$$

She does not know where to start as the expression has brackets, multiple terms and uses all arithmetic operations.

Fatima should simplify the expression step-by-step as follows:

Step 1: Simplify all terms within the brackets as much as possible

$$2(3x^2 + 5z^3 + \frac{3z^4}{z} + 2zx^3) - (2y + 5x^2 - y)$$

$$= 2(3x^2 + 5z^3 + \frac{3z^{\not{4}3}}{\not{z}} + 2zx^3) - (2y + 5x^2 - y)$$

$$= 2(3x^2 + (5z^3 + 3z^3) + 2zx^3) - ((2y - y) + 5x^2)$$

$$= 2(3x^2 + 8z^3 + 2zx^3) - (y + 5x^2)$$

Step 2: Since 2 precedes the first brackets, multiply each term within the first brackets with 2.

$$2(3x^2 + 8z^3 + 2zx^3) - (y + 5x^2)$$

$$= (6x^2 + 16z^3 + 4zx^3) - (y + 5x^2)$$

Step 3: Check the sign that precedes each bracket. The first set of brackets are preceded by a plus sign, therefore the sign of all the terms within these brackets will remain the same when the brackets are removed. However, the second set of brackets is preceded by a minus sign. The signs of all terms within these brackets will change when the brackets are removed.

Therefore,

$(6x^2 + 16z^3 + 4zx^3) - (y + 5x^2)$

$= 6x^2 + 16z^3 + 4zx^3 - y - 5x^2$

$= x^2 + 16z^3 + 4zx^3 - y$

OVER TO YOU

Activity 4: Using brackets

Simplify the following algebraic expression.

$$1 - (-2x^2y + 3z^3) + 2(x^3z + (6xy \times 3z) - 2zx^3 + x^3z) - \left(2xy + \frac{5x^2}{xy} - xy\right) - 5y$$

2.2 Solving equations

Equations: an introduction

We have so far expressed mathematical relationships as algebraic expressions. These expressions can be presented as equations to arrive at solutions to problems.

> *An algebraic equation is a mathematical statement that depicts the equality of two expressions.*

Let's look at Gemma's monthly earnings and expenditure (Case study 1) again. Gemma's earnings were stated in an algebraic expression as:

$100x - y - 20$

If in January, after all her expenditure during the month, the balance remaining with Gemma was $500, this relationship can be expressed as:

$100x - y - 20 = 500$

This expression is an **equation** with two variables x and y.

An equation has two sides. These two sides should always be equal for the equation to stay true or valid. If any change of variables or constants is done on one side of the equation, the same change should also be done on the other side of the equation. For example, if a number is added or subtracted on one side, the same number should be added or subtracted from the other side. Similarly, if the terms on one side are multiplied or divided by a number, the same number should multiply or divide the terms on the other side also.

Algebraic equations have unknown values that can only be known after the equation is solved. An equation can be solved by separating unknown quantities (variables) from known quantities (constants) or by reorganising the equation.

Algebraic equations can be of various types. The focus of this section is on learning how to solve the following types of equations.

1 Single variable **linear equations**
2 **Simultaneous linear equations** in two variables
3 **Quadratic equations**

Single variable linear equations

A linear equation is an equation that has one or more variables. A linear equation has two properties:

- The maximum power of each variable is one.
- When plotted on a graph, the equation gives a straight line.

> *A single variable linear equation has only one unknown quantity or variable.*

CASE STUDY 1 (CONTINUED)

Gemma's monthly earnings and expenditure

In February, Gemma's monthly expenses excluding the magazine subscription were 60% of her monthly wages. At the end of the month, she had $740 remaining with her after all her monthly expenditure. She decides to put this amount in her savings bank account. You want to know how many days did Gemma work in February.

Gemma's earnings in February can be expressed as an algebraic equation:

$100x - 0.6(100x) - 20 = 740$

This is a linear equation in one variable x, where x represents the number of days for which Gemma worked during the month of February. This equation can be solved by determining the value of x.

The steps are:

Step 1: Simplify both sides of the equation by clearing brackets and combining like terms.

$100x - 0.6(100x) - 20 = 740$
$100x - 60x - 20 = 740$

Step 2: Separate the variable from constants by bringing all the terms with the variable on one side of the equation and moving all the constants on the other side.

This movement of algebraic terms from one side to the other is known as **transposition.** Transposition helps to separate the unknown variable on one side so that a solution can be found.

Now simplify each side as explained here:

$100x - 60x - 20 = 740$

$100x - 60x - 20 + 20 = 740 + 20$

$100x - 60x = 740 + 20$

$100x - 60x = 760$

All the terms containing the unknown x are kept on left hand side of the equation and all the known values are moved to the right-hand side.

Note carefully, that during transposition, the sign of each term changes when it is moved to the other side of the equation.

Step 3: Bring the coefficient of the variable down to one.

- If the coefficient of the variable is an integer, divide both sides by that integer.
- If the coefficient of the variable is a fraction, multiply both sides of the equation by the reciprocal of the fraction.

In this example, the coefficient of the variable is an integer '40'. Therefore, divide both sides of the equation by 40.

$$100x - 60x = 760$$

$$40x = 760$$

$$\frac{40x}{40} = \frac{760}{40}$$

$$x = 19$$

The solution to the equation is the value that equates the variable when its coefficient as 1.

In this example $x = 19$.

Therefore, Gemma worked for 19 days in February.

Steps for solving a single variable linear equation

Step 1: Simplify both sides of the equation by clearing brackets and combining like terms.

Step 2: Get all terms with the variable on one side of the equation and move all constants on the other side to simplify each side.

Step 3: Bring the coefficient of the variable down to one. The solution to the equation is the value that equates the variable when its coefficient as 1.

REVISION on the go

OVER TO YOU

Activity 5: Solving single variable linear equations

Solve the following linear equations

1 $2u - 6 = -14$

2 $3(x + 6) + 3x = 3(-9 - x)$

3 $\frac{2t}{t + 3} - 2 = \frac{3}{t - 10}$

Solving simultaneous linear equations in two variables

As we have seen, a simple linear equation in one variable can be easily solved. However, if a linear equation has more than one variable, there can be more than one solution for each variable. Take this example:

$$x + y = 5$$

This is a linear equation in two variables x and y, which can take multiple values such as (5,0), (4,1), (3,2), (2,3), (1,4), (0,5) and so on. To solve this kind of an equation, we require another equation that has a relationship between the two unknown variables. The two equations are then used at the same time or simultaneously to arrive at the correct pair of values for x and y. The two equations used to arrive at the answer to the two unknown variables are called **simultaneous equations**.

Simultaneous equations must be consistent and independent.

CASE STUDY 2

Solving simultaneous equations

Marie has two equations to solve. She has to find the values of x and y. The equations are:

$5x + 2y = 144$

$2x + 2y = 72$

The equations Marie is solving are simultaneous linear equations in two variables x and y.

Methods of solving simultaneous equations

There are two methods of solving such equations:

1 Elimination method
2 Substitution method

Elimination method

In the elimination method, one of the variables is removed from the two equations. This is based on the premise, that if the relationship shown in the two simultaneous equations is true, then any equation that is derived from this relationship is also true.

Elimination method: one variable in the two equations has the same coefficient

This can be easily done if one of the variables has the same coefficient regardless of sign of the coefficient. The following steps need to be performed:

Step 1: See whether one of the two unknowns has the same coefficient. If yes, identify it. In these equations y has the same coefficient '2' in both equations.

Step 2: If the variable that has the same coefficient has opposite signs, then add the two equations. However, if the variable with the same coefficients has the same sign, then subtract the two equations from each other. This step will eliminate the variable.

In Marie's case as the variable y in both equations has a + sign, the second equation is subtracted from the first equation:

$$+5x + 2y = 144$$

$$\underline{-2x + -2y = -72}$$

$$+3x + 0 = 72$$

Step 3: The equation becomes a single variable linear equation, which can be solved for the remaining variable, that is, x.

$$+3x + 0 = 72$$

$$3x = 72$$

$$\frac{3x}{3} = \frac{72}{3}$$

$$x = 24$$

Step 4: Replace x with its value in either of the equations to find the value of y.

$2x + 2y = 72$

$2(24) + 2y = 72$

$48 + 2y = 72$

$48 + 2y - 48 = 72 - 48$

$2y = 24$

$y = 12$

Therefore, $x = 24$ and $y = 12$.

You can verify your answer by substituting the values you have calculated in either of the equations.

$5x + 2y = 144$

$5(24) + 2(12) = 144$

$144 = 144$

Elimination method: none of the two variables in the two equations has the same coefficient

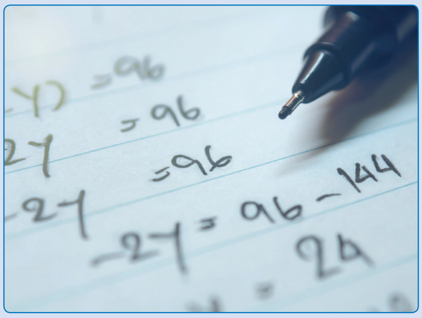
CASE STUDY 2 (CONTINUED)

Simultaneous equations

Marie now has two equations in which neither of the two variables has the same coefficients. She wants to solve these equations to get the values for x and y. The equations are:

$6u + 2v = 56$

$5u - v = 76$



In these equations, the coefficients of both variables are different. Such equations can be solved by manoeuvring one or both equations so that at one variable gets same coefficients in both equations. This will help in eliminating the variable.

Step 1: Identify the variable whose coefficient can easily be made the same in both equations.

The coefficient of 'v' in equation 2 can be made same as that in equation 1 by simply multiplying each term on both sides of equation 2 by 2 to get a new equation (equation 3).

$6u + 2v = 56$ .... **Equation 1**

$5u - v = 76$ .... **Equation 2**

$2 \times 5u - 2 \times v = 76 \times 2$

$10u - 2v = 152$ .... **Equation 3**

Step 2: Now consider equation 1 and equation 3.

$6u + 2v = 56$ .... **Equation 1**

$10u - 2v = 152$ .... **Equation 3**

Variable v in both equations has the same coefficient but with opposite signs. We can therefore add the two equations to eliminate v.

$$\begin{array}{r} 6u + 2v = 56 \\ \underline{10u - 2v = 152} \\ 16u + 0 = 208 \end{array}$$

Step 3: It now becomes a single variable linear equation that can be solved for the remaining variable, that is, u.

$16u + 0 = 208$

$16u = 208$

$\frac{16u}{16} = \frac{208}{16}$

$u = 13$

Step 4: Substitute the value of u in either of the equations and find the value of v.

$6u + 2v = 56$

$6(13) + 2v = 56$

$78 + 2v = 56$

$78 + 2v - 78 = 56 - 78$

$2v = -22$

$v = -11$

Therefore, $u = 13$ and $v = -11$

You can verify your answer by substituting the values you have calculated in either of the equations.

OVER TO YOU

Activity 6: Simultaneous equations

Using the elimination method, solve the following equation by eliminating u rather than v.

$6u + 2v = 56$

$5u - v = 76$

Hint: you will need to manoeuvre both equations.

Elimination method: variables appear as denominators of fractions

Using elimination method can be difficult when variables appear as denominators of fractions. For example,

$\frac{4}{x} - \frac{1}{y} = 9$ .... **Equation 1**

$\frac{1}{x} + \frac{1}{y} = 11$ .... **Equation 2**

These can be solved as follows:

Step 1: Assume $\frac{1}{x} = r$ and $\frac{1}{y} = s$ and rewrite the equations as:

$4r - s = 9$

$r + s = 11$

Step 2: In these equations s has the same coefficient '1' in both equations. As s has opposite signs in the two equations, add the two equations to eliminate s.

$4r - s = 9$
$r + s = 11$
$5r = 20$

Step 3: It now becomes a single variable linear equation which can be solved for the remaining variable, that is, r.

$5r = 20$
$r = \frac{20}{5} = 4$

Step 4: Substitute the value of r in either of the equations and find the value of s.

$r + s = 11$

$4 + s = 11$

$4 + s - 4 = 11 - 4$

$s = 7$

Since $\frac{1}{x} = r$ *and* $\frac{1}{y} = s$, *the answer is* $\frac{1}{x} = 4$ *and* $\frac{1}{y} = 7$

You can verify your answer by substituting the values you have calculated in either of the equations.

OVER TO YOU

Activity 7: Simultaneous equations

Find the values of x and y in the following equations.

1. $3x - y = 2$
   $2x + y = 13$
2. $x + y = 7$
   $3x + 2y = 11$
3. $5x - 5y = 15$
   $2x + 6y = 30$
4. $2x + 6y + 4 = 0$
   $7x + 8y = 12$

Substitution method

As the name suggests, with substitution method, the value of one variable is stated in terms of the second variable. The value derived for the first variable is then used in the second equation. This eliminates the first variable in the second equation and leaves a linear equation in single variable. The second variable in this equation can be solved easily.

Now Marie (Case study 2) solves the second set of equations by using the substitution method. The steps are:

Step 1: Use the first equation to express the value of u in terms of v. Please note that you can use any one of the two equations to get the value of any one of the two variables.

$6u + 2v = 56$

$6u + 2v - 2v = 56 - 2v$

$6u = 56 - 2v$

$\frac{6u}{6} = \frac{56 - 2v}{6}$

$u = \frac{56 - 2v}{6} = \frac{\cancel{2}(28 - v)}{\cancel{6}3} = \frac{28 - v}{3}$

Step 2: Replace u with $\frac{28 - v}{3}$ in the second equation.

$5\left(\frac{28 - v}{3}\right) - v = 76$

$\frac{140 - 5v}{3} - v = 76$

$\frac{140 - 5v - 3v}{3} = 76$

$\frac{140 - 8v}{3} = 76$

$140 - 8v = 76 \times 3$

$140 - 8v = 228$

$140 - 8v - 140 = 228 - 140$

$-8v = 88$

$v = \frac{88}{-8} = -11$

Step 3: Replace v with –11 in the first equation to get the value of u.

$6u + 2v = 56$

$6u + 2(-11) = 56$

$6u - 22 = 56$

$6u - 22 + 22 = 56 + 22$

$6u = 78$

$u = 13$

Therefore, $u = 13$, $v = -11$.

OVER TO YOU

Activity 8: Simultaneous equations: substitution method

Using the substitution method solve Marie's first set of simultaneous equations.

$5x - 2y = 144$

$2x + 2y = 72$

Formulating simultaneous equations from given situations

Let's look at Gemma's monthly earnings and expenditure (Case study 1). We can see that real-life situations and problems can be modelled as simultaneous equations. Unknown values, that may be important for decision-making, can easily be obtained by solving these equations.

CASE STUDY 3

Formulating simultaneous equations

There is a new fast food restaurant in town. The restaurant charges different prices for an adult's meal and a child's meal. Sue and Mike take their two children to the restaurant and order four meals. They pay $55. Ahmed takes his three children to the same restaurant and orders four meals. He pays $48. We will now:

1 Formulate the simultaneous equations
2 Calculate the price of an adult's meal and a child's meal.

The number of meals in both cases is the same but since the restaurant charges different prices for an adult's meal and a child's meal, the total amount of money paid by Sue and Mike is different from the amount paid by Ahmed. There are two unknowns in this situation: (1) price of an adult's meal and (2) the price of a child's meal. The value of these two unknowns can be determined with simultaneous equations.

Let,

x = price of an adult's meal

y = price of a child's meal

Equation 1: Sue and Mike with their two children implies 2 adult's meals and 2 children's meals, with a total payment of $55. This can be expressed as equation:

$2x + 2y = \$55$

Equation 2: Ahmed with his three children implies 1 adult's meal and 3 children's meals, with a total payment of $48. This can be expressed as equation:

$x + 3y = \$48$

Solving these two simultaneous equations will tell us the price of an adult's meal x and a child's meal y. The equations can be solved using either elimination or substitution method.

$2x + 2y = \$56$

$x + 3y = \$48$

We can give x the same coefficient in both equations by multiplying the second equation with 2. Remember that to have a balanced equation, you must multiply all elements of the equation with 2. This gives a new equation:

$2 \times x + 2 \times 3y = 2 \times 48$

$2x + 6y = 96$

After this step, the two equations can now be solved as:

$2x + 2y = 56$

$$\begin{array}{r} -2x + -6y = -96 \\ \hline 0 - 4y = -40 \end{array}$$

$y = \frac{40}{4} = \$10$

Next, substitute $y = 10$ in the first equation.

$2x + 2y = \$56$

$2x + 2(10) = 56$

$2x + 20 = 56$

$2x + 20 - 20 = 56 - 20$

$2x = 36$

$x = \frac{36}{2}$

$x = \$18$

Thus, the price of an adult's meal is $18 and that of a child's meal is $10.

OVER TO YOU

Activity 9: Simultaneous equations

Using the substitution method, find the price of an adult's meal and a child's meal.

$2x + 2y = \$56$

$x + 3y = \$48$

Where,

x = *price of an adult's meal*

y = *price of a child's meal*

Solving quadratic equations

Linear equations, including simultaneous equations, have variables to the power of 1. Quadratic equations are formulated from quadratic expressions. A quadratic expression is an algebraic expression in which the highest power of a variable is 2.

Example $6x^2 - 4x$

Example $t^2 - t - 2$

Quadratic expressions can also exist in equations in which the power of a variable is a multiple of 2. For example $2x^4 + 3x^2 + 7$ is a quadratic expression of an algebraic relationship in x^2 because this equation can also be expressed as $2(x^2)^2 + 3(x^2) + 7$. If it is assumed that $x^2 = y$, then the expression can be written as $2y^2 + 3y + 7$ which is a normal quadratic expression.

A quadratic equation is formulated by equating a quadratic expression to a similar expression or a number. In the simplest form, quadratic equations take the form of $ax^2 + bx + c = 0$, where a, b and c are numbers and $a \neq 0$

For example,

$3x^2 - 4x + 1 = 0$

or

$u^2 - u = 5$

There is no definitive solution of a quadratic equation because usually the solution to a quadratic equation has two possible values for the unknown. Linear equations, when plotted on a graph, give a straight line, quadratic equations, when plotted on a graph, result in a curve.

Methods of solving quadratic equations

This section focuses on solving quadratic equations using the following two methods:

1 The factorisation method
2 The quadratic formulae

The factorisation method

The basic premise of the factorisation method is that if the product of two terms is zero, then either one or both of the terms must also be equal to zero. Therefore, if $xy = 0$ then either $x = 0$ or $y = 0$ or both are equal to 0. Similarly, if $x(x - 1) = 0$ then either $x = 0$ or $(x - 1) = 0$ or both are equal to zero. This method is applicable when the right-hand side of the equation can be made equal to 0 and the left-hand side can be expressed as factors.

Let's solve a quadratic equation $2x^2 - 7x = 7$ by using the factorisation method.

This would require you to be able to present the equation as the product of two numbers and equate it to 0. The steps are:

Step 1: Collect all the terms of the quadratic equation to left side and make the equation equal to 0.

$2x^2 - 5x = 7$

Therefore, $2x^2 - 5x - 7 = 0$

Step 2: Factorise the left-hand side of the equation. This involves presenting the left-hand side as a product of two algebraic expressions. This can be done by using the factors of the product of coefficients of a and c in the equation $ax^2 + bx + c = 0$ and organising them such that their sum is equal to the coefficient of the middle term b.

In example $2x^2 - 5x - 7 = 0$, $a = 2$ and $c = -7$.

The product of these numbers = –14.

The factors of –14 are –7 and 2 and

$-7 + 2 = -5$ which is the coefficient of b.

Therefore left-hand side of the equation can be factorised as follows:

$2x^2 - 5x - 7$

$= 2x^2 - 7x + 2x - 7$

$= x(2x - 7) + 1(2x - 7)$

$= (x + 1)(2x - 7)$

The equation $2x^2 - 5x - 7 = 0$ can therefore be written as $(x + 1)(2x - 7) = 0$

Step 3: Take each factor after factorisation and equate it to 0 and work the two values of x.

Therefore,

$(x + 1) = 0$

$x + 1 = 0$

$x = -1$

and

$(2x - 7) = 0$

$2x - 7 = 0$

$2x = 7$

$x = \frac{7}{2} = 3\frac{1}{2}$

Therefore, $x = -1$ or $x = 3\frac{1}{2}$

OVER TO YOU

Activity 10: Quadratic equations: factorisation method

Solve the following quadratic equations

1. $2x^2 - 5x - 3 = 0$
2. $4t^2 - 9t + 5 = 0$
3. $x^2 + 6x + 8 = 0$
4. $3y^2 + y = 2$

CASE STUDY 4

Solving quadratic equations

Asha wants to construct a fence around her rectangular garden in a way that the length of the garden is 4 feet more than its width. The area of the garden is 60 square feet. She wants to know what would be the length and the width of her garden.

The area of a rectangle is equal to the product of its length and width, both of which are unknown values in this example.

Let the length of the garden be x. As width is 4 feet less than the length, it can be written as $(x - 4)$. Algebraically, this equation can be presented as:

$x(x - 4) = 60$

This can be solved using the factorisation method as follows:

$x(x - 4) = 60$

$x^2 - 4x = 60$

$x^2 - 4x - 60 = 0$

$x^2 + 6x - 10x - 60 = 0$

$x(x + 6) - 10(x + 6) = 0$

$(x + 6)(x - 10) = 0$

Therefore

$(x + 6) = 0$

$x + 6 = 0$

$x = -6$

or

$(x - 10) = 0$

$x - 10 = 0$

$x = 10$

Since the length or width of the garden cannot be a negative value, the length = 10 feet and the width = 10 − 4 = 6 feet.

The quadratic formula

You may not find it easy to factorise all quadratic expressions. In many cases, we need to use the quadratic formula to solve the equations.

The quadratic equation expressed as $ax^2 + bx + c = 0$ can be solved with the formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

CASE STUDY 4 CONTINUED

Solving quadratic equations

Asha drops her earlier plan. She now decides to build a fence that is 6 feet more than its width. The area of the garden is 60 square feet.

In this instance of case study 4, you will not easily to find the values of length and width using the factorisation method. The equation can be expressed as:

$x(x - 6) = 60$

$x^2 - 6x - 60 = 0$

The equation can be solved using the formula $x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

Where,

$a = 1,$

$b = -6,$

$c = -60$

Therefore,

$$x = \frac{-(-6) \pm \sqrt{(-6)^2 - 4(1)(-60)}}{2(1)}$$

$$= \frac{6 \pm \sqrt{36 + 240}}{2}$$

$$= \frac{6 \pm 16.6}{2}$$

The value of x can be either $\dfrac{6 + 16.6}{2} = 11.3$ or $\dfrac{6 - 16.6}{2} = -5.3$

Since the length of the garden cannot be negative, therefore the length $(x) = 11.3$ feet, and the width is 11.3 – 6 = 5.3 feet

OVER TO YOU

Activity 11: Quadratic equations: formula method

Asha has decided to construct a garden that is square shaped. Calculate the length of the garden if the area of the garden is 60 square feet.

OVER TO YOU

Activity 12: Quadratic equations: formula method

Solve the following quadratic equations.

1 $x^2 - x - 7 = 0$

2 $4x^2 - 2x - 3 = 0$

3 $9x^2 + 12x - 4 = 0$

4 $x^2 = 2x + 1$

Deriving the equation of a straight line

This section focuses on understanding and deriving the equation of a straight line or a linear equation. Before discussing the equation of a straight line, let us refresh our understanding of the framework of graphs.

> *A graph is a diagrammatic illustration of relationship between variables.*

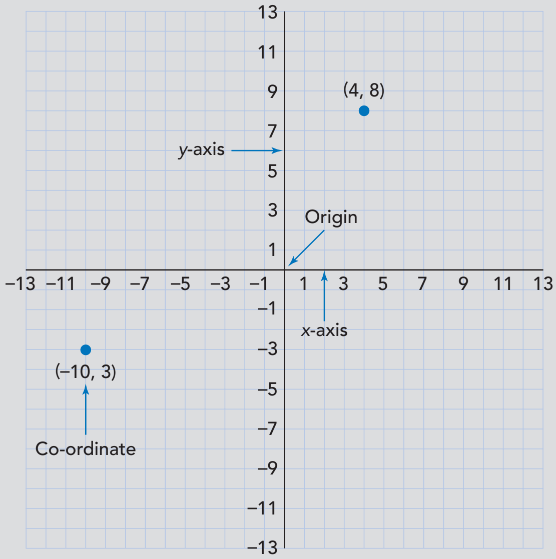


Figure 2: Basic framework of a graph

REVISION on the go

Figure 2 shows the framework of a typical graph.

- A graph has two axes: the horizontal straight line or x-axis and vertical straight line or y-axis.
- The x-axis depicts the values that the independent variable can take and the y-axis depicts the values that the dependent variable can take.
- The various points on the graphs are referenced by its **co-ordinates**. A co-ordinate is a pair of values of x and y variables that indicate a specific position on the graph. These are presented as (x, y); the first value indicates the position on the x-axis and the second value indicates the

position on the y-axis. For example, (4, 8) in Figure 2 refers to a point at the intersection of 4 on the x-axis and 8 on the y-axis.

> *A dependent variable is a variable whose value depends on an independent variable.*

A linear equation when plotted on a graph always gives a straight line. For plotting a straight line on a graph, we will need a minimum of three co-ordinate points derived from the linear equation that needs to be plotted. To do so, choose three values of x (independent variable) and by using the linear equation find the corresponding values of y (dependent variable). The three x and y co-ordinates when plotted on a graph and joined will give a straight line.

Consider an example. We want to plot the following linear equation on a graph:

$y = 2x + 1$

Choose three values of x and find the corresponding values of y.

x	$y = 2x + 1$
0	2(0) + 1 = 1
2	2(2) + 1 = 5
4	2(4) + 1 = 9

This gives three co-ordinates for the graph (0, 1), (2, 5) and (4, 9). These co-ordinates when plotted and joined give the graph for $y = 2x + 1$. This is illustrated in Figure 3.

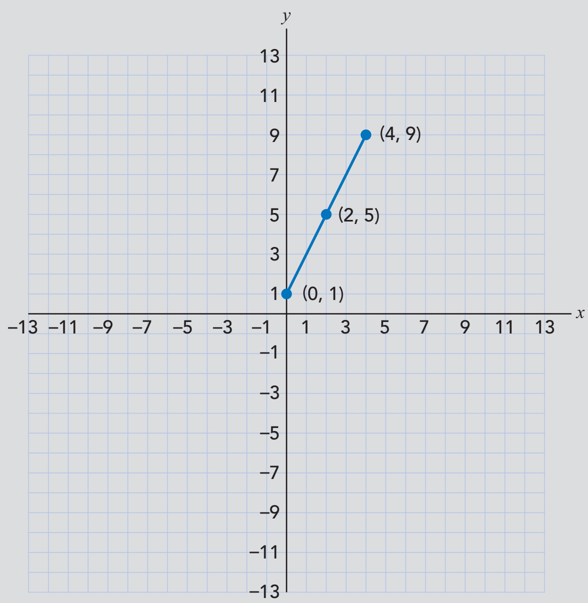


Figure 3: Graph of $y = 2x + 1$

Therefore, a linear equation is an equation of a straight line. The straight line can be extended and for any value of x, the value of y can be found out.

OVER TO YOU

Activity 13: Plotting an equation

Given equation $y - 3x = 2$

1 Plot the graph for the equation.

2 State the shape of the graph.

A linear equation or the equation of a straight line takes the form of $y = mx + c$

where,

x is the **independent variable** and y is the **dependent variable**

m represents the **gradient** or slope of the line. For example, the slope of $y = 2x + 1$ is 2.

c represents the point where the straight line crosses the *y*-axis. It is also known as the **y-intercept**.

In Figure 3, the straight line crosses the *y*-axis at $y = 1$. Therefore, the *y*-intercept is 1.

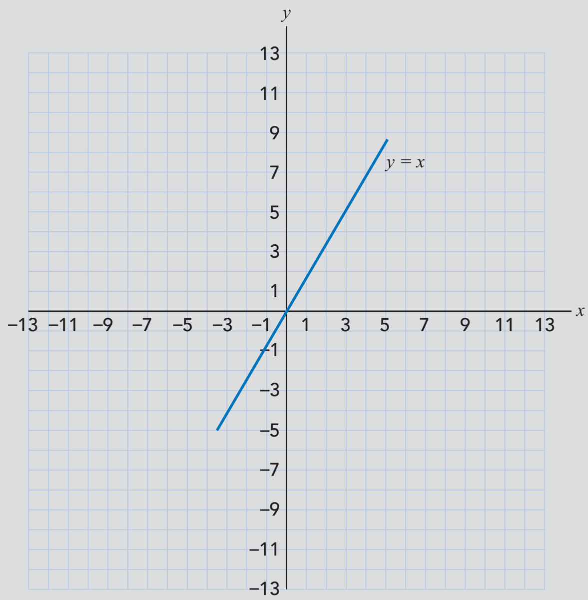


Figure 4: Graph of a straight line passing through the origin

The equation of a straight line that passes through the origin is $y = mx$ as the y-intercept in this case is 0. This is shown in Figure 4.

Equation of a straight line

A linear equation or the equation of a straight line takes the form of $y = mx + c$

The equation of a straight line that passes through the origin is $y = mx$

Ascertaining the gradient and y-intercept of an equation for a straight line

The formula of gradient can be derived by re-arranging the linear equation $y = mx + c$ in terms of m.

$y = mx + c$

$y - c = mx$

$mx = y - c$

$m = \frac{(y - c)}{x}$

We can also find the gradient of a straight line by using the formula:

Gradient $= \frac{\text{Change in } x}{\text{Change in } y}$

$m = \frac{(y_2 - y_1)}{(x_2 - x_1)}$

Where,

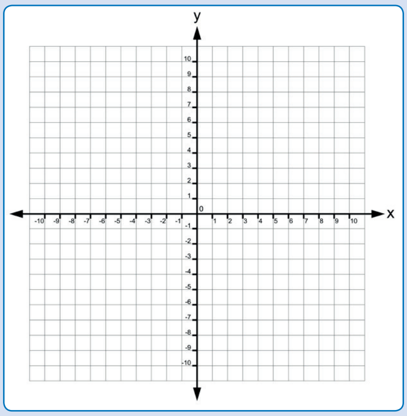
y_1 and y_2 are two points on the y-axis and x_1 and x_2 are corresponding points on the x-axis.

CASE STUDY 5

Jiang's assignment on the equation of a straight line

Jiang' is doing an assignment on equations of straight lines.

He wants to determine the gradient and y-intercept in the equation $2y - 5x = 8$



The equation $2y - 5x = 8$ is not currently in its general form and should be put in the form $y = mx + c$ to determine the gradient and y-intercept. For this, Jiang should transpose all terms except y to the right-hand side of equation

$2y - 5x = 8$

$2y = 8 + 5x$

$y = \frac{8 + 5x}{2}$

$y = 4 + \frac{5}{2}x$

$y = \frac{5}{2}x + 4$

Therefore, the gradient $= \frac{5}{2}$ and y–intercept = 4.

Determining gradient (m)

$$m = \frac{(y - c)}{x}$$

$$m = \frac{\text{Change in } x}{\text{Change in } y} = \frac{(y_2 - y_1)}{(x_2 - x_1)}$$

REVISION on the go

OVER TO YOU

Activity 14: Gradient and *y*-intercept of an equation of a straight line

Find the gradient and *y*-intercept in each equation.

1. $x + 3y + 2 = 0$
2. $2x + y = 4$
3. $3y = 9 - 5x$
4. $\frac{1}{2}x - 2 = y$

Determining the equation of a straight line when its gradient and *y*-intercept are known

It is easy to derive the equation of a straight line when both gradient and *y*-intercept are given. In cases like these, put the given values of gradient and *y*–intercept in the general form of the equation of a straight line, that is, $y = mx + c$.

Consider an example. The gradient of a straight line is $-\frac{1}{2}$ and the *y*–intercept is $\frac{3}{5}$. The equation for a straight line that can be derived from the available information is:

$$y = -\frac{1}{2}x + \frac{3}{5}$$

OVER TO YOU

Activity 15: Deriving the equation of a straight line

Derive the equation of the straight line for each of the following:

1. Gradient = $\frac{1}{6}$ and *y*-intercept = –5
2. $y = 1$ and $m = 0.7$

Determining the equation of a straight line given its gradient and one point lying on it

CASE STUDY 5 (CONTINUED)

Jiang's assignment on the equation of a straight line

Jiang wants to derive another equation of a straight line, for which he only knows that the co-ordinates of one point lying on the line are (4,3) and the gradient of the line is $\frac{1}{3}$.

In this case, Jiang knows the gradient of the line but does not know its y-intercept. This equation can be easily derived in two steps.

Step 1: In the equation of a straight line put $m = \frac{1}{3}$ and use the co-ordinates of the given point for x and y. This will help derive c, the y-intercept.

$$y = mx + c$$

$$x = 4,\ y = 3,\ m = \frac{1}{3}$$

Therefore,

$$3 = \frac{1}{3} \times 4 + c$$

$$3 - \frac{4}{3} = c$$

$$c = \frac{5}{3}$$

Step 2: Insert the values of the gradient (m) and y-intercept (c) in the general equation of a straight line to get the required equation.

$$y = mx + c$$

$$y = \frac{1}{3}x + \frac{5}{3}$$

Determining the equation of a straight line when two points lying on it are known

The equation of the straight line that passes through two given points can be written as:

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}$$

Alternatively, it can be written as:

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}$$

$$y - y_1 = \left(\frac{x - x_1}{x_2 - x_1}\right)(y_2 - y_1)$$

$$y - y_1 = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)(x - x_1)$$

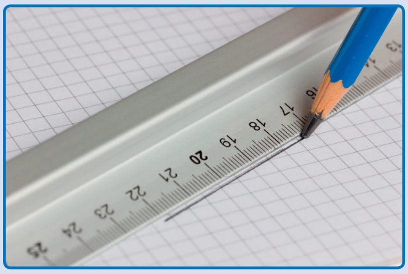
$$y - y_1 = m(x - x_1)$$

y_1 and y_2 are two points on the y-axis and x_1 and x_2 are corresponding points on the x-axis.

CASE STUDY 5 (CONTINUED)

Jiang's assignment on the equation of a straight line

Jiang knows that he can draw a straight line by connecting two points. He also knows the two points lying on the straight line are (2,4) and (3,5). He wants to derive the equation from these points, but he does not know the gradient or y-intercept of the line.



This equation of a straight line can be derived from the co-ordinates of two points lying on it by using the equation:

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}$$

In this case,

$y_1 = 4$, $y_2 = 5$, $x_1 = 2$, $x_2 = 3$

Therefore,

$$\frac{y - 4}{5 - 4} = \frac{x - 2}{3 - 2}$$

$$\frac{y - 4}{1} = \frac{x - 2}{1}$$

$$y - 4 = x - 2$$

$$y = x - 2 + 4$$

$$y = x + 2$$

Alternatively, the equation can be derived by using the equation $y - y_1 = m(x - x_1)$.

In this case, the steps are:

- Determine the slope m by using the formula $m = \frac{(y_2 - y_1)}{(x_2 - x_1)}$
- Use the co-ordinates of one of the points in $y - y_1 = m(x - x_1)$ to derive the equation.

In this example $m = \frac{5 - 4}{3 - 2} = 1$

Therefore, using (2,4), the equation will be:

$$y - y_1 = m(x - x_1)$$

$$y - 4 = 1(x - 2)$$

$$y - 4 = x - 2$$

$$y = x - 2 + 4$$

$$y = x + 2$$

Deriving the equation of a straight line given two points lying on it

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}$$

or

$$y - y_1 = m(x - x_1)$$

y_1 and y_2 are two points on the y-axis and x_1 and x_2 are corresponding points on the x-axis.

REVISION on the go

OVER TO YOU

Activity 16: Deriving the equation of a straight line

1 Derive the equation of the line which:

a passes through (2, 3) and has a gradient of –3

b passes through (1, 6) and (–2, 0)

c passes through (1, –1) and has a gradient of –1

d passes through (–2, 3) and (4, 0).

2 Write down the y-intercept in each case.

NEED TO KNOW

An algebraic term may be expressed as simply a letter or a combination of a number and a letter.

An algebraic expression is a collection of algebraic terms used to express a relationship. It uses integers, constants, variables and arithmetic operators to express a relationship.

Only like terms in an algebraic expression can be added or subtracted.

Like terms can be multiplied together by raising the product to the power of the number of times the term appears.

Two expressions with the same base can be multiplied even if they have different exponents or powers.

A term raised to the power of another term is denoted with the base raised to the power of the product of the two exponents.

Unlike addition and subtraction, terms that are NOT like terms can also be multiplied and the product is obtained by simply writing the terms being multiplied next to each other.

If one or more terms have coefficients greater than 1, then coefficients are multiplied and the resulting value is written before the product of the letters.

Algebraic terms can be easily cancelled out while dividing.

Two expressions with the same base but different exponents can be divided by each other by simply taking the difference between the exponents.

A term having a negative exponent can be expressed as its reciprocal with a positive exponent.

It is a good practice to simplify all terms within the brackets as much as possible. This makes the algebraic expression more manageable. There are rules that apply to brackets in algebra.

An algebraic equation is a mathematical statement that depicts the equality of two expressions. The two sides should always be equal for the equation to stay true or valid. If any change of variables or constants is done on one side of the equation, the same change should also be done on the other side of the equation.

Algebraic equations have unknown values that can only be known after the equation is solved.

A linear equation is an equation which has one or more variables.

Two properties of a linear equation are:

- the maximum power of each variable is one;
- when plotted on a graph, the equation gives a straight line.

The two equations used to arrive at the answer to the two unknown variables are called simultaneous equations.

There are two methods of solving simultaneous equations: Elimination method and substitution method.

Under elimination method, one of the variables is removed from the two equations. This is based on the premise, that if the relationship shown in the two simultaneous equations is true, then any equation that is derived from this relationship is also true.

Under substitution method, the value of one variable is stated in terms of the second variable. The value derived for the first variable is then used in the second equation. This eliminates the first variable in the second equation and leaves a linear equation in single variable.

Quadratic equations are formulated from quadratic expressions. A quadratic expression is an algebraic expression in which the highest power of a variable is 2.

Quadratic equations can be solved using the factorisation method. The basic premise of the factorisation method is that if the product of two terms is zero, then either one or both of the terms must also be equal to zero.

At times it is not easy to factorise all quadratic expressions. In many cases we need to use the quadratic formula to solve the equations.

A linear equation when plotted on a graph always gives a straight line and takes the form of $y = mx + c$

In cases when both gradient and *y*-intercept are given, put the given values of gradient and *y*–intercept in the general form of the equation of a straight line, that is, $y = mx + c$.

The equation of the straight line that passes through two given points can be written as:

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}$$

READING LIST

ABE (2017). Introduction to Quantitative Methods. ABE. ISBN 978-1-911550-13-6

Huettenmueller, R. (2011). Algebra DeMYSTiFieD. McGraw Hill Education, 2nd Ed. ISBN 978-0071743617

Larson, R.E. and Hodgkins, A.V. (2011). College Algebra with Applications for Business and Life Sciences. Brooks/Cole; 2nd International edition. ISBN 978-0547052694

READING RESOURCES

BBC KS3 Bitesize (2017). Algebra. BBC website, available at http://www.bbc.co.uk/bitesize/ks3/maths/algebra/

Chapter 3

Business Statistics

Introduction

Statistics is all about data. There are two important aspects related to data. The first is that data is not always in the form of numbers. The second is that for many problems, data is obtained only from a sample of a large population, and then it is used to extract meaningful information for making decisions. This chapter will focus on three topics. First, it will examine the various sources from which data can be obtained, the types of data, and its use in business. Next it will explore sampling techniques involved in data collection. Finally, it will discuss the classification and tabulation of collected data.

Learning outcome

After completing this chapter, you will be able to:

3 Discuss the process of gathering business and management data (Weighting 25%)

Assessment criteria

3.1 Explain main sources, types and uses of data relevant for business and management information

3.2 Evaluate alternative methods of sampling and measurement scales used in context of business information

3.3 Classify and tabulate statistical data

Data in statistics

Statistics is a specialised branch of mathematics consisting of methods that are used to systematically collect, organise and analyse data with the aim to get meaningful information that can be used for decision-making.

There are two branches of statistics:

Descriptive statistics	*Consists of methods for organising and summarising information* (Weiss, 1999)
Inferential statistics	*Consists of methods for drawing and measuring the reliability of conclusions about population based on the information obtained from a sample of population* (Weiss, 1999)

The branches are tightly integrated. Inferential statistics draws upon the methods of descriptive statistics. Data is the core of statistics.

There are important factors to consider before you undertake a statistical study:

- state the objective clearly
- define the scope of research
- the required accuracy of data.

The stages of a statistical study are depicted in Figure 1. Collection and classification of data is discussed in Chapter 3. Analysis of data and presentation of results will be discussed in Chapter 4.

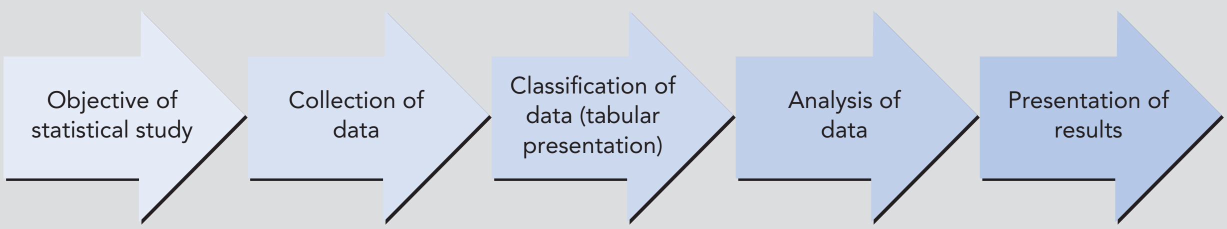


Figure 1: Stages of statistical study

CASE STUDY 1

David and Fatima's research study

David is a university student. He learns from a newspaper article titled "The age of fast fashion" that 80% of university students buy at least one item of casual clothing every week. According to the report, 70% students consider price as the most important factor when they shop for clothes, and only 30% of students consider style and design as a more important factor than price. He shares this information with his classmate, Fatima. Both agree to undertake similar research on their own university campus.

David and Fatima need information on how students in their university buy items of casual clothing and the importance they give to factors such as price, style, and design while making a purchase. For this they will need to collect data.

Data plays an important role in the functioning of a society. Individuals, business organisations and governments receive data on a regular basis and use it for decision-making. There are many definitions of data.

The Oxford dictionary defines data as:

> *Facts collected for reference or analysis.*

The Economic Commission for Europe and United Nations (UNECE) gives a more elaborate definition:

> *Data is the physical representation of information in a manner suitable for communication, interpretation or processing by human beings or by automatic means.*

Data is facts that can be either numbers or attributes. When data is collected for the first time, it is raw and unorganised. Decisions cannot be taken with **raw data**. Therefore, it is important to organise and process raw data into information before it can be used for decision making.

Number of sales invoices for t-shirts today is 14	14 (data) is a number.
Different coloured polo t-shirts were sold	Colour (data) is an attribute.

Table 1: Examples of data

REVISION on the go

Let's look at Table 1, showing sales of t-shirts in a shop. It has examples of data that can be processed to obtain information. On analysis, the shop owner learns that:

- 14 customers purchased polo t-shirts today. He learns this only after he has counted the sales invoices;

- blue coloured t-shirts have maximum sales. He learns this only after he has organised a large number of sales invoices for a certain period of time.

The shop owner can use this information to make a decision on what colours of t-shirts to stock.

OVER TO YOU

Activity 1: Data and information

Make a list of data that can be gathered from a monthly mobile phone bill. What useful information can be gained after this data is analysed?

Sources of data

Let's look at David and Fatima's research study (in case study 1). The initial source of data for David was a newspaper. Data obtained from such pre-existing published or unpublished sources is called **secondary data**. Other sources of secondary data include journals, government reports, television news, and articles on the internet.

Now David and Fatima want to do similar research in their university. Therefore, they decide to collect data directly from the university students who are now the subject of their statistical study. This original data that they will be collecting is called **primary data**. Sources of primary data in a business context are surveys of customers, employees and suppliers with the aid of questionnaires, focus groups, personal interviews, and direct observations.

The differences between secondary and primary data are presented in Table 2.

Basis of difference	Secondary data	Primary data
Origin	Already existing due to someone's earlier effort	Collected from original field work by the researcher
Relevance	May not relate directly to the research objective	Relates directly to the research objective
Effort required in collection	Easier to collect, with lesser time, money and energy	Consume more time, money and energy compared to secondary data

Table 2: Differences between secondary and primary data

REVISION on the go

Questionnaires are widely used for a variety of surveys. These include market research, government censuses and opinion polls. A focus group is a specific type of group interview used for understanding customer attitudes and perceptions, and also testing new product concepts.

The international statistical institute defines a questionnaire as:

> *"A group or sequence of questions designed to elicit information upon a subject, or sequence of subjects, from an informant."*

The Organisation of Economic Co-operation and Development (OCED) defines a focus group as:

> *"An interviewing technique whereby respondents are interviewed in a group setting."*

OVER TO YOU

Activity 2: Primary and secondary data

Classify these examples as primary data or secondary data.

1. **John checked the bills and counted the number of customers who had placed an order for a chicken burger at the counter.**
2. **Ahmed likes gaming apps. He checks a website to find out the list of apps that have received a rating of 5 stars.**
3. **Nancy takes the heights of her classmates at school and makes a record of their measurements.**
4. **Li Wei checks a website to find out the value of Chinese exports to the United Kingdom in 2015 and 2016.**

Secondary and primary data

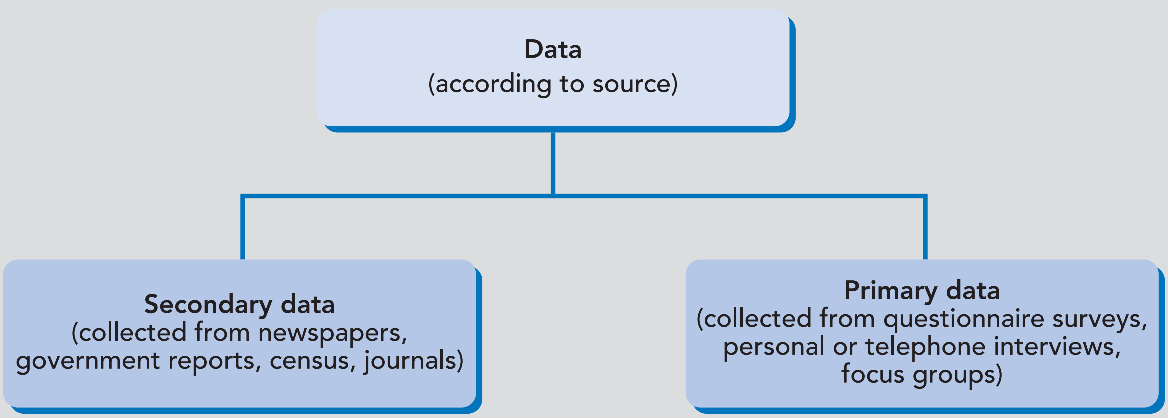


Figure 2: Types of data according to source

REVISION on the go

3.1 Difference between types of data

Types of data

Secondary and primary data can be either quantitative or qualitative. Quantitative data is numeric observations, and qualitative data is non-numeric (or categorical) observations.

Quantitative data is further classified into discrete data and continuous data.

An example of quantitative discrete data is:

The number of t-shirts purchased by a customer at a shop: 3

The values for quantitative data can only be integers.

"Shoe size is a notable exception of discrete data that can have a fractional value. For example, a shoe size of 4½ is discrete quantitative data."

An example of quantitative continuous data is:

The time (in minutes and seconds) spent by one customer in the shop.
This can take many values: 15.2 or 15.23 or 15.235 or 15.2354 and so on.

It is evident that all fractional values of time in quantitative continuous data make sense. The same is not true for discrete data; the value 2.35 clothes has no meaning.

An example of qualitative data is:

The colour of t-shirts purchased by one customer: blue

The differences between discrete and continuous data are presented in Table 3.

Basis of difference	Discrete data	Continuous data
Generated by	Counting	Measuring
Value	Can take only finite values between stated limits	Can take Infinite values between two stated limits
Effort required in collection	Values are almost always integers; cannot be fractions.	Can be fractions and decimals.

Table 3: Differences between discrete and continuous data

Variables

Table 4 gives data for the number and colour of t-shirts bought by six customers. In this table, the value '3' for customer A will have a meaning only if it relates to the characteristic 'number of t-shirts'. Similarly, the colour 'blue' has meaning only when it relates to the characteristic 'colour of the t-shirts'. These characteristics are called variables because their data values keep changing for each subject.

A researcher can collect data for a single or multiple variables.

The Australian Bureau of Statistics defines a variable as:

> *A variable is any characteristics, number, or quantity that can be measured or counted. A variable may also be called a data item.*

Customers	**Variable: Number of t-shirts bought**	**Variable: Colours of t-shirts bought**
Customer A	3	Blue
Customer B	2	Red
Customer C	5	Yellow
Customer D	2	Grey
Customer E	1	Green
Customer F	4	Purple

Table 4: Quantitative and qualitative variables

- Variables for which numeric data is collected are called quantitative variables; for example, number of t-shirts.
- Variables for which non-numeric data is collected are called qualitative variables; for example, colours of t-shirts bought.

Data values collected for a variable from several customers when presented together are called a data set. In this case, two data sets are obtained.

- The numeric data set {3, 2, 5, 2, 1, 4} for the quantitative discrete variable 'number of t-shirts bought'.
- The categorical data set {red, blue, yellow, grey, green, purple} is for the 'colour of t-shirts bought', which is a qualitative variable.

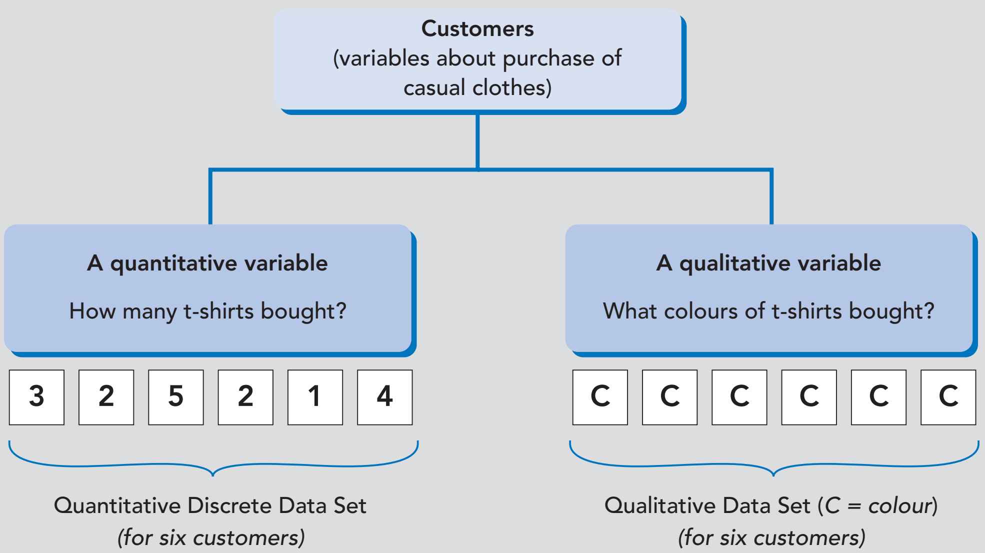
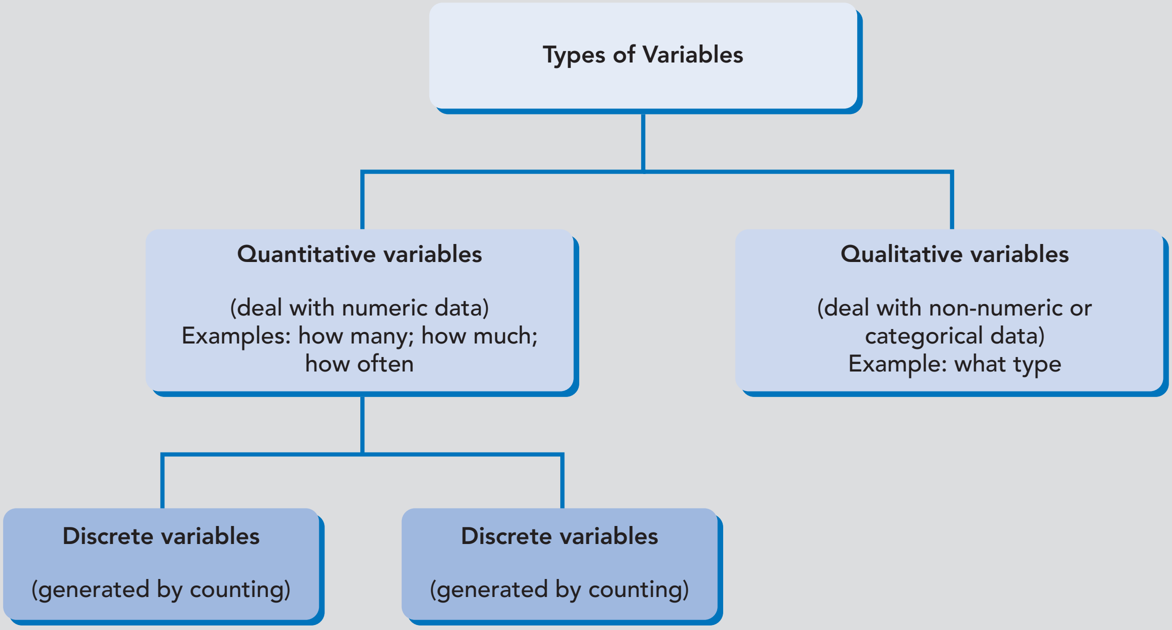


Figure 3: Data, variables and data sets

Types of variables



Quantitative variables + Qualitative variables → Complete information on the subject

Adapted from: Australian Bureau of Statistics website

Figure 4: Types of variables

REVISION on the go

OVER TO YOU

Activity 3: Types of data

Alia wants to buy a fashionable case for her new mobile phone. She visits an online shop. State which type of data Alia will get from each of the questions below.

1. **What different types of materials are cases available in?**
2. **What is the price for each type of case?**

OVER TO YOU

Activity 4: Discrete and continuous data

The owner of a retail outlet wants to find out the number of items sold last month. Is this data discrete or continuous?

OVER TO YOU

Activity 5: Discrete and continuous data

Classify each of the following as discrete or continuous data.

1 Height of a student
2 Number of books
3 Time
4 Temperature
5 Shoe size

OVER TO YOU

Activity 6:

Refer to Case study 1. If you're studying with others, form groups of two students in your classroom.

1 List down examples of 10 questions that will help David and Fatima to collect qualitative and quantitative data for getting the complete picture of their population's shopping habits and preferences for clothes.
2 Classify your examples of quantitative data into discrete and continuous.
3 Also mention the method that you will use for data collection and state the reason for choosing your method.

Use of data in business

The 21st century business environment is fast paced and full of uncertainty. Globalisation and the arrival of the internet has made the business environment much more dynamic. Innovative start-up firms are challenging firms that have been leaders in their industry for a long time. Proactive decisions are needed now more than ever. Smart managers understand this new reality and use the large volume of data generated each day:

- spot market trends, and exploit them to their advantage;
- analyse costs and expenses, and manage them to get better profitability and shareholder return;
- understand customer behaviour and expectations, and devise ways to cultivate a large base of loyal customers that can generate sales revenue on a continuous basis, for example, by developing new or enhanced products and identifying ways to improve customer satisfaction.

The aim is better decision-making and superior organisational performance. Some information is also shared with the investors and government. Figure 5 depicts the various kinds of data generated in a business organisation (in rectangles) and some examples of its use (in the central circle).

Since the arrival of computers, it takes less human effort to process large volume of data. However, interpretation and communication of data requires people with special skills.

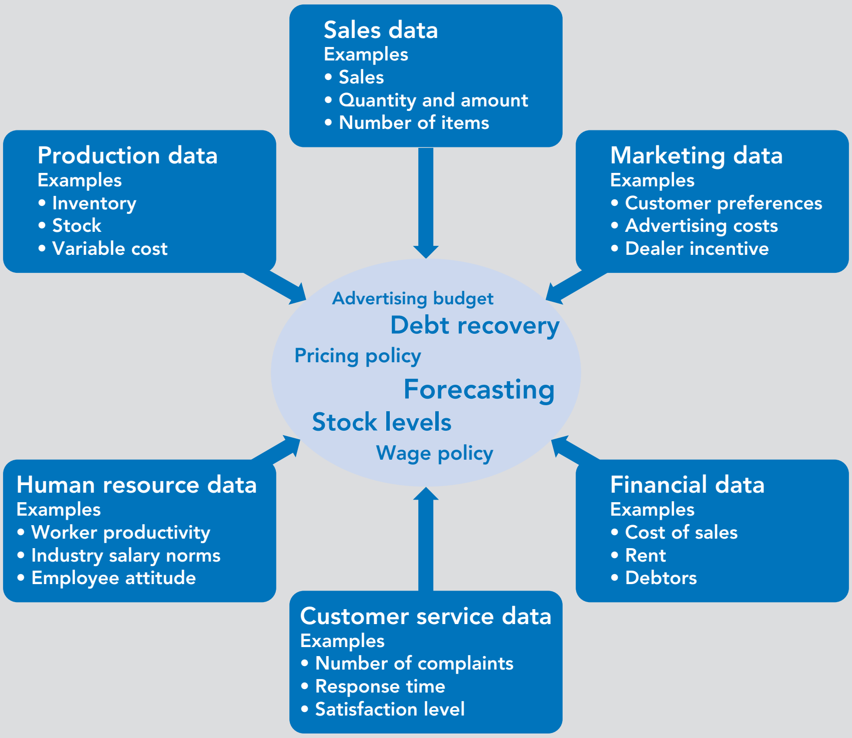


Figure 5: Examples of use of data in business

REVISION on the go

3.2 Sampling methods and measurement scales

Sampling techniques

CASE STUDY 1 CONTINUED

David and Fatima's research study

Let's look at case study 1 again to examine a new piece of information.

There are 6000 students in the university in which David and Fatima plan to conduct their research. It is not possible for David and Fatima to collect data from all the university students. They'll need to work out what to do next.

The subject of a statistical study is called the population. This population can be people, objects or institutions. In David and Fatima's research study, the population is 6000 students studying in their university. Gathering data from a large population like this is both time-consuming and costly. A practical way to collect data is to select a subset of this population. This subset is called a **sample**. Data collected from a sample is organised and analysed with the application of statistical methods with the aim to draw conclusions about the population.

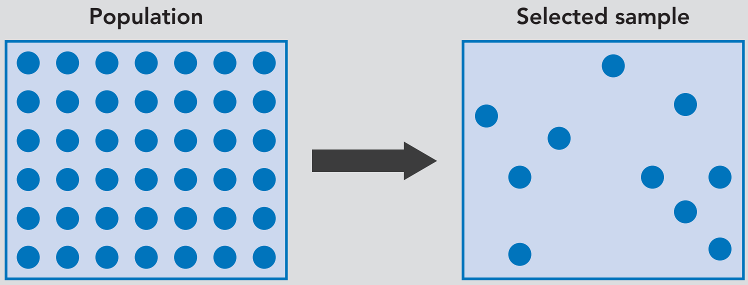


Figure 6: Population versus sample

You can see the sampling process in Figure 7. The key terms are explained in Table 5.

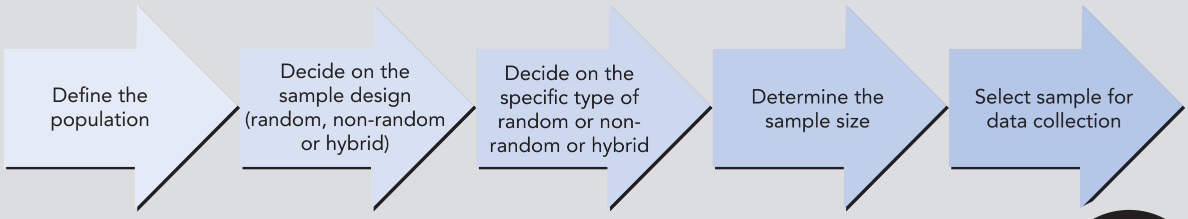


Figure 7: Sampling process

Key terms	
Census	This is the act of collecting data from every person or about every object present in a national geography. A government of a country conducts this kind of large-scale survey.
Population	This is the complete set of people or objects that is the subject of statistical study. Example: Students studying in a university.
Parameter	This is the numerical property of a population.
Sampling unit	This is the most fundamental building block of a sample. This can be either an individual or an object that can be counted, measured or observed for a statistical study.
Sample	This is an appropriately sized subset of the population from which data is collected to draw true and unbiased conclusions about the population under study. Example: A sample of students on the campus.
Frame	This is a published source on population that is the subject of statistical study. A frame may not list the complete population, but is often used for selecting a sample. Examples: A directory of schools; or a telephone directory of listed households in a city.

Table 5: Key terms in the sampling process

OVER TO YOU

Activity 7: Census

1 Find out when was the last time a census was conducted in your country.
2 List five key items of information that are produced by a census.

It is important to define the population and sample before data collection is initiated. The researcher should decide a sample carefully. According to Newbold, Carlson and Thorne (2012: 7), the key factors to consider in making sample choices are:

- objectives of the study;
- nature and definition of population;
- research design considerations;
- availability of resources;
- ethical and legal considerations.

The size of the sample is an important consideration in research. Sample size should be optimum, i.e. neither too small nor too large. It is determined by two factors:

- Cost – a sample size that is larger than required can be expensive.
- Precision – a sample size, if smaller than required, may fail to give true picture of the population.

There are three sampling techniques:

1 random sampling (also called probability sampling);
2 non-random sampling (also called non-probability sampling);
3 mixed-method sampling (hybrid sampling) combines both random and non-random sampling.

The key differences between random and non-random sampling is presented in Table 6.

Basis of difference	Random sampling	Non-random sampling
Approach	Random selection of subjects	Deliberate, intentional selection of subjects
Chance given to every unit in the population	Equal chance, therefore this is known as probability sampling	Unequal chance, therefore this is known as non-probability sampling
Cost	Expensive	Relatively less expensive
Time	Time consuming	Speedy
Preparation in data collection	More	Relatively less
Reliability	Sample is more likely to be representative of the population	Sample is less likely to be representative of the population
Scope	May require access to the whole population	There is no need to access the entire population

Table 6: Differences between random and non-random sampling

Random and non-random sampling can be applied with various techniques presented in Table 7.

Random sampling techniques	Non-random sampling techniques
• Simple random • Systematic • Stratified random • Cluster (or Area)	• Quota • Convenience • Judgment • Snowball

Table 7: Techniques of random and non-random sampling

REVISION on the go

Random sampling techniques

CASE STUDY 2

Chen's research

Chen has been asked to carry out research on the companies based in a local industrial park. The objective of the research is to understand the methods of recruitment that these companies prefer to use for attracting entry-level employees, for example, job advertisements, recruitment consultants, university campus visits, referrals, etc.

Chen decides to approach companies located in an industrial park in his town, and gets the directory that lists all companies that operate in the industrial park. The directory has names, addresses and telephone numbers of 40 companies. He decides to choose a sample of companies from this frame by using a random sampling technique.

Simple random sampling

Refer to Chen's research (case study 2). The frame, which in this case, is the directory of the industrial park, has names of 40 organisations.

1. Apple	11. Experian	21. Johnson	31. Pembroke
2. Accenture	12. Gems	22. JP Morgan	32. Paragon
3. Aveva	13. Glencore	23. Joules Group	33. Paegent
4. Barratt	14. Genus	24. Keller Group	34. Sicoro
5. Boohoo	15. Greencoat	25. Martinco	35. Summit
6. Barclays	16. Greencore	26. Microfocus	36. Stride
7. Burberry	17. Halfords	27. Mitsubishi	37. Topps
8. Ceres	18. IDOX	28. Nokia	38. Troy
9. Centrica	19. JD Sports	29. Nasstar	39. Warring
10. Dods Group	20. Jupiter	30. Oxford Bio	40. Xerox

Table 8: Types of random and non-random sampling

REVISION on the go

Chen aims to collect data from eight organisations with the aid of a questionnaire. He decides to use random sampling technique that gives all subjects (in this case organisations) in the frame equal probability of being included in the sample. The steps involved in selecting the sample are:

1 List all members in the frame, numbering them from 1 to 40.

2 Use the random number function on a scientific calculator or spreadsheet software to generate a random number between 1 and 40.

3 Select the subject whose number matches the random number and add it to the sample list.

4 Repeat this, seven more times. Ignore any number that is repeated, and generate a new number.

For example, if the random numbers generated are: 25, 6, 35, 40, 20, 11, 29, 31, then the subject selected as sample from the frame are:

Martinco, Barclays, Summit, Xerox, Jupiter, Experian, Nasstar and Pembroke.

Systematic sampling

This is a sampling technique in which the sample is made of subjects that are placed at equal intervals. If the required sample size is 8 and the given frame has 40 subjects, then the selection of 8 subjects will be determined by:

1 the first sampling unit selected randomly;

2 an appropriate interval ***k*** determined by the formula ***k* = *N*/*n*** (where *N* is the size of the frame, and *n* is the sample size).

In this case, the frame size is 40 and the sample size is 8, therefore, the interval is 5. If number 2 is the starting number, then the numbers selected for sample will be:

2, 7, 12, 17, 22, 27, 32 and 37

This implies that the subjects that will make up the sample include:

Accenture, Burberry, Gems, Halfords, JP Morgan, Mitsubishi, Paragon and Topps.

Stratified random sampling

A population or frame may look homogenous but, in reality, is heterogeneous. This means that the population or frame is made up of subgroups called **strata**. The members within each subgroup are similar, but there is some degree of variability between each subgroup.

Consider an example of a population composed of individuals in the age group 15 to 45 years. This population is the subject of statistical study on how people buy casual clothes and how often do they buy. On the surface, this population looks homogenous because everyone buys casual clothes. However in reality, this population may be heterogeneous. This implies that the younger subgroup of 15–25 years of age may be buying a lot more casual clothes and much more frequently than members of population in 25–35 years age group and 35–45 years age group. Therefore, the use of simple random sampling may give a **sampling error** if the selected sample has a large number of sampling units from one subgroup and a very small number from other two subgroups. The information obtained from such a sample may not offer a true picture of the population. This sampling error can be avoided by using stratified sampling technique.

Let's look at Chen's research (Case study 2) again. Chen realises that his frame has two subgroups:

- 12 organisations that are small and medium enterprises (SMEs) with less than 250 employees;
- 28 organisations that are large organisations with more than 250 employees.

Chen decides to use stratified random sampling for greater accuracy. He aims to collect data for understanding:

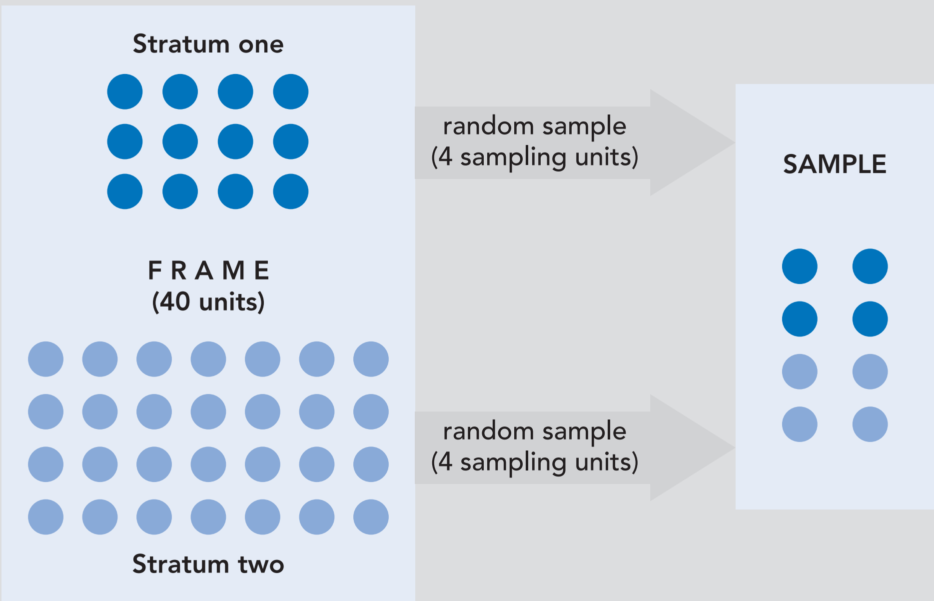
1. the methods that organisations prefer for entry-level recruitment;
2. whether SMEs and large organisations adopt similar methods.

Chen aims to select a sample of eight subjects (organisations), with four sampling units from each sub-group.

The steps involved in selecting the sample are:

1. Divide the population or the frame into two subgroups or **strata** according to a specific attribute. In this example, the attribute is size of the organisation.
2. Organise the two subgroups and number each subject in each subgroup. For example, for stratum one comprising of 12 SMEs, number the subjects from 1 to 12, and for stratum two, comprising of 28 large organisations, number the subjects from 1 to 28.
3. Apply simple random sampling technique on each stratum. In this context, generate four random numbers between 1 and 12 for stratum one, and, generate four random numbers between 1 and 28 for stratum two.
4. Select the organisations that correspond to the random numbers for each stratum.

Stratified sampling process is depicted in Figure 8.



Stratum One: 12 small and medium (SME) organisations
Stratum Two: 28 large organisations

Figure 8: Stratified random sampling

REVISION on the go

The biggest advantage of a stratified sampling technique is that it has a greater chance of representing the population than both simple random and systematic random sampling. This reduces the sampling error. However, the disadvantages are that it is costly and time consuming.

In statistics, a stratum (plural strata) is a subgroup within a population or a frame that has properties that are distinct from other subgroups.

REVISION on the go

Cluster sampling

This sampling technique is also called area sampling. The population is divided into clusters that are natural miniatures of the population. Each cluster has all kinds of members.

Let's look at Chen's research (case study 2) again. Instead of considering one industrial park in his town as his population, Chen redefines the population for his study as all the industrial parks at various locations within his region. He further defines each industrial park as one cluster. For example, if there are six industrial parks in Chen's region, then the population is made up of six clusters. He further defines his sample size as any ten organisations from any two clusters, i.e. five sampling units from each cluster. He then decides to apply a simple random sampling technique for selecting his sample.

The steps involved in selecting the sample are:

1 List all clusters in the population, numbering them from 1 to 6. For each cluster (industrial park) create a numbered list of all its member organisations.

2 Use the simple random sampling technique to generate two random numbers between 1 and 6. Select the industrial parks whose number matches the random numbers.

3 Apply the simple random sampling technique again to randomly select five organisations from each of the two selected clusters.

4 Collect data from the selected subjects within the two clusters.

The difference between stratified sampling and cluster sampling is depicted in Figure 9.

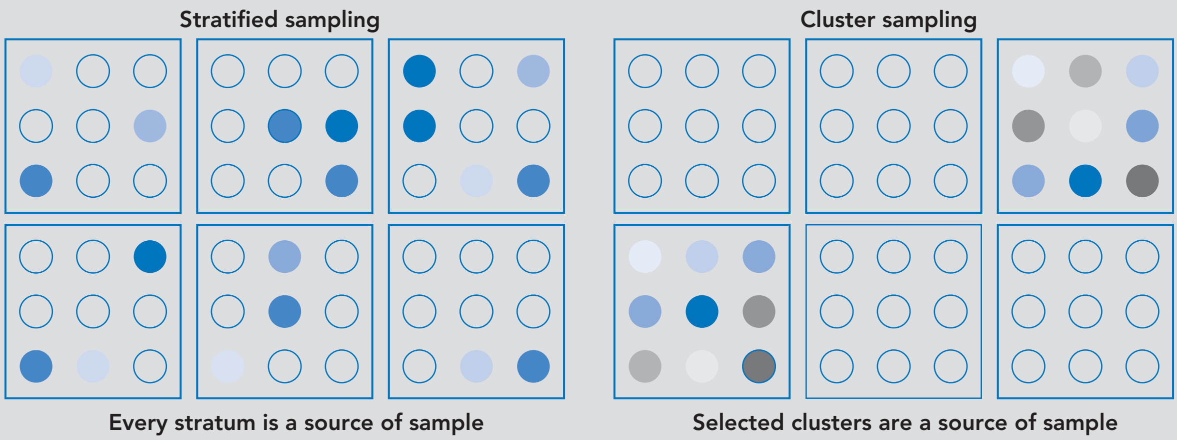


Figure 9: Stratified vs. Cluster sampling

REVISION on the go

OVER TO YOU

Activity 8: Sampling techniques

Bilal is conducting research through which he wants to understand whether male and female school teachers adopt different teaching styles. His population consists of 12 schools in his small town. What sampling technique has been used by Bilal for the following two approaches?

1. **He decides to select a sample with equal numbers of male and female teachers from all the 12 schools.**
2. **From the 12 schools, he decides to interview all teachers of only two schools, A and B. School A is a boys' school with all male teachers. School B is girls' school with all female teachers.**

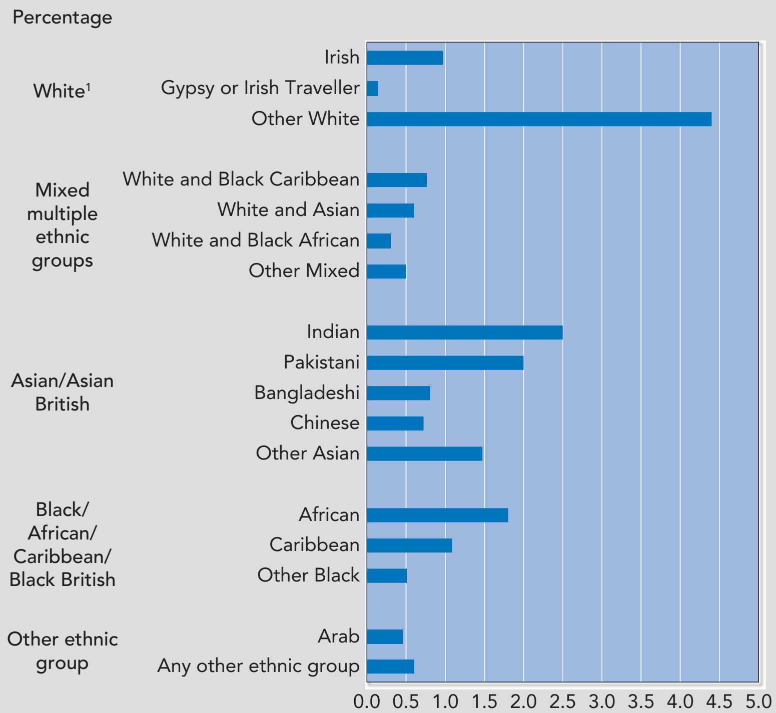
Non-random sampling techniques

Unlike random sampling techniques, non-random sampling techniques do not rely on chance. The researcher conducting the statistical study chooses the sample deliberately. For this reason, non-random sampling techniques are called non-probability techniques.

Quota sampling

In this approach to sampling, the population is divided into strata. However, instead of random selection of sampling units from each stratum, the sample is selected in a non-random manner. A quota is fixed for each stratum. The quota may be decided according to the proportion of the strata within the population. The researcher aims to collect data from each stratum according to the assigned quota.

Quota sampling is a non-random variation of stratified sampling. It is less expensive than stratified sampling because it is easier to manage. Quota sampling is widely used for opinion polls.



Source: Office of National Statistics website

Figure 10: Ethnic strata in England and Wales

Let's look at Figure 10. The population is divided according to ethnic groups. The researcher assigns the quota as: White (100 sampling units); Asian British (45 sampling units); and Black British (35 sampling units).

The data, when it is analysed, gives information about each subgroup, and the whole population.

Convenience sampling

As the name suggests, the researcher's convenience determines what sampling units are selected. According to Black (2009), convenience is based on three factors:

- availability
- proximity or accessibility
- willingness of the sampling unit to participate.

An example of convenience sampling based on easy accessibility is selection of households that have no pet dogs and are located on the main-streets.

Judgement sampling

Professional researchers often use their earlier experience and judgement to select a sample. The main aim of using non-random judgement sampling is to save time and money. However, if the researcher's judgement is not balanced, there is a risk of **bias**.

Two similar but independent statistical studies that use judgement sampling can produce very dissimilar pictures of the population.

Snowball sampling

This is an inexpensive way of sampling in which one sampling unit leads the researcher to the next sampling unit through reference. The premise of this technique is that referrals are likely to be both, accessible and willing to participate in the survey.

Let's look at the David and Fatima's research study (case study 1) again. If David and Fatima use snowball sampling technique, then their sample will be selected as:

1. Interview the first student who is willing to respond to their questionnaire.
2. After collecting data, ask the student to suggest names of friends who buys casual clothing, and would be willing to answer their questions.
3. Make a contact with the referrals and collect data.
4. Ask the interviewees for more referrals and repeat the process till sample size is achieved.

Sampling error and bias

Sampling error is the consequence of selecting a sample whose measurement offers a result that does not reflect the true picture of the population characteristic being studied. Sampling error is not a deliberate mistake. It takes place because in many instances, selection is approximate rather than exact. Sampling error is measurable. However, you cannot completely eliminate it, but can certainly minimise it.

Bias is a consequence of adopting a practice that distorts data and therefore leads to an untrue picture of the population. Bias is a non-random, systematic error. Examples include: interviewer bias, error in recording, questionnaires that are not well-designed, measurements that are not carefully taken, respondents who do not answer truthfully.

Sampling approaches and techniques

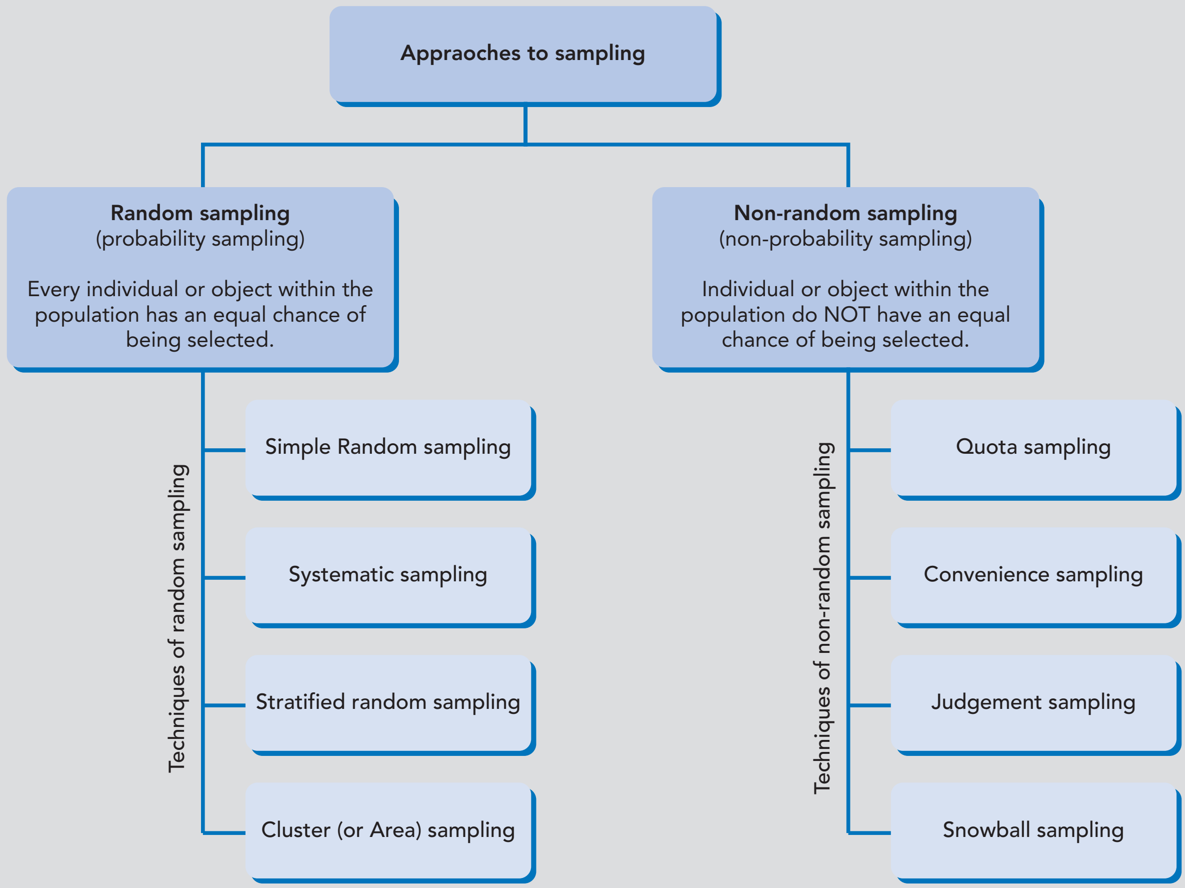


Figure 11: Sampling approaches and techniques

Measurement scales

Now look at Figure 12. There are four different questions, which, when answered by a respondent, help in taking measurements.

Question One	Question Two
Which course are you doing?	How would you rank three subjects in order of your preference? *Marketing, Human resources and Business statistics*
Response: 1 Mathematics 2 Computer Science 3 **Business Management** 4 Chemistry 5 History	Response: 1 Business statistics 2 Marketing 3 Human resource

Question Three

How would you rate your last lecture on business statistics on a 5-point scale, starting from lowest (1) to highest (5)?

1	2	3	4	**5**

Question Four

How many times did you get textbooks issued from the university library in the last three terms?

Response:

- Term 1 3
- Term 2 9
- Term 3 6

Figure 12: Measurement scales

REVISION on the go

Let's understand the significance of how each question is being measured.

- **Question One:** Let us suppose that the respondent has chosen the value 3, i.e. business management. This value only signifies that the respondent is a student of the business management course. The values 1, 2, 3, 4 and 5 in this question are not indicative of the importance of the course. The values do not imply that business management course has a lower or higher ranking than other courses.

  This kind of data measurement scale used for classification is called nominal level data measurement.

- **Question Two:** Let us suppose that the respondent ranks the subjects as shown in Figure 12. By ranking the three subjects, the respondent is only indicating his preference. This kind of measurement clearly shows that the respondent prefers business statistics to marketing, and marketing to human resources. However, the ranking does not tell us whether the degree of the respondent's preference for any two subjects is equal.

  The data measurement scale that is used for ranking elements is termed as ordinal level data measurement.

- **Question Three:** You can see that the scale that the respondent will use to rate the business statistics lecture in this example has equal intervals. The data is numeric. This implies that ranking of lecture is measurable with this scale.

  This scale is termed as interval level data measurement.

  An important feature of interval level data measurement is that zero is just one of the many intervals. A good example of this measurement is the temperature scale with which readings are taken at every degree change in temperature. Zero degree is not the lowest value on this scale because there are negative and positive temperature intervals on either side.

- **Question Four:** It is understood that no student can get books issued from the library less than zero times. Therefore, the value zero has significance in this measurement scale. It is a fixed point that gives meaning to other data values. A value can also be expressed as a ratio of another value and therefore data can be interpreted meaningfully. For example, in term three, the library was used twice as much as in term one for getting textbooks.

  This scale of data measurement is ratio level data measurement.

OVER TO YOU

Activity 9

What level of data measurement is gender classification? Explain your answer.

3.3 Classification and tabulation of data

Data collected from research is raw data. Meaningful information and conclusions can be drawn from this data only after it is:

1. classified
2. presented in the form of tables or charts and diagrams
3. analysed.

This chapter will focus on classification and presentation of data using tables. Presentation of data using charts and diagrams, and analysis of data will be discussed in Chapter 4.

Three main types of classification are:

- simple classification resulting in simple tables;
- complex classification resulting in two-way tables;
- multi-dimensional classification resulting in manifold tables.

> *Classification is the act of arranging items in groups according to their similarity.*

Rules for tabulation

Tables involve the classification and presentation of data in rows and columns. We call this process tabulation. If you create a table properly, you will get meaningful information. There are some basic rules for a good table structure:

- Use a reasonable size that is easily comprehended at first look.
- Make sure it has a purpose, communicated with a brief title.
- Don't overload it with data.
- There should be a visible arrangement, for example, alphabetical or chronological order.
- Columns should have a brief title or a number.

- There should be a relationship between columns.
- Group similar items.
- Units of measurement should be explicit, for example currency or weight.
- Present column and row totals if they are important information.
- Footnotes and references should be a part of table, wherever applicable.

Let's look at the David and Fatima's research study (case study 1) again. David and Fatima collected raw data from a mix of male and female respondents. They arranged the questionnaires filled with responses in alphabetical order. The numeric data against each name represents "the number of clothes bought" by the respondents for their own use in the last three months.

Ahmed	Male	7
Alice	Female	7
Asha	Female	6
Ben	Male	1
Elizabeth	Female	14
George	Male	5
Jodie	Female	7
John	Male	10
Julie	Female	9
Kartik	Male	5
Li Na	Female	12
Nancy	Female	11
Paul	Male	4
Roger	Male	6
Samir	Male	5
Stella	Female	8
Sue	Female	8
Zoya	Female	8

Simple tables and two-way tables

This chapter discusses only simple and complex (two-way) classification. Multidimensional classification is not within the scope of this syllabus.

Simple classification

David and Fatima use simple classification to construct a table. The classification is based on the number of clothes that were bought by male and female students. The data is presented in Table 8.

Gender	Number of clothes bought
Male students	43
Female students	90
Total	133

Table 8: A table showing simple classification

REVISION on the go

A simple table is a result of simple, one-way classification of data in rows and columns that gives information about one variable only.

Complex classification

David and Fatima have also collected data on the type of casual clothes that the student bought. Two of the clothing items are t-shirts and jeans. You can see the raw data for number of t-shirts (column 3) and jeans (column 4) bought in the last three months below.

Ahmed	Male	5	2
Alice	Female	4	3
Asha	Female	4	2
Ben	Male	1	0
Elizabeth	Female	10	4
George	Male	4	1
Jodie	Female	5	2
John	Male	7	3
Julie	Female	6	3
Kartik	Male	3	2
Li Na	Female	10	2
Nancy	Female	6	5
Paul	Male	2	2
Roger	Male	5	1
Samir	Male	4	1
Stella	Female	6	2
Sue	Female	4	4
Zoya	Female	3	5

David and Fatima use complex classification to construct a two-way table that gives information on the number of t-shirts and jeans bought (variable 1) by male and female students (variable 2). You can see the data in Table 9.

Gender	Number of clothes bought by type		Total
	T-shirts	Jeans	
Male students	31	12	43
Female students	58	32	90
Total	89	44	133

Table 9: A two-way table

"A two-way table is the result of complex classification of data in rows and columns that gives information about two inter-related variables."

OVER TO YOU

Activity 10: Two-way tables

Jiao and Waheeda are conducting research on the gaming market. They decide to randomly interview 20 young male adults and 20 young female adults. They find that 13 males play games only on their mobile phones and 7 males play games only on their gaming consoles, while 15 females play games only on their mobile phones and 5 females play games only on their gaming consoles.

Prepare a two-way table to present this data.

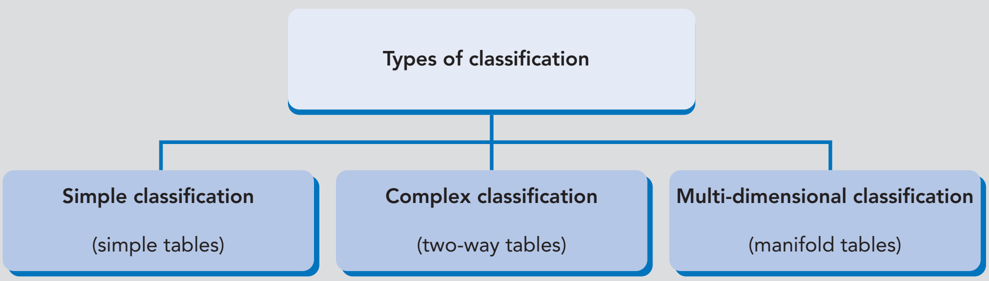


Figure 13: Types of classification

Frequency distribution

Frequency distribution is tabular classification of data for large data sets. A frequency distribution table shows the number of times (or frequency) a variable takes each of its data values.

Simple frequency distribution

Let's look at the data that David and Fatima collected on the number of shopping trips that each respondent made in the last three months for buying clothes.

Ahmed	4	Kartik	4
Alice	6	Li Na	4
Asha	7	Nancy	4
Ben	4	Paul	3
Elizabeth	4	Roger	3
George	5	Samir	4
Jodie	6	Stella	5
John	4	Sue	5
Julie	4	Zoya	6

Use of tally marks

Tally marks or hash marks are often used to record the number of instances (frequency) a variable takes a value. In this example, the variable is "number of shopping trips". The method of using tally marks is:

1. Identify the lowest value and the highest value.
2. Create appropriate intervals for the entire range.
3. For each interval, draw one tally mark for every instance of value that relates to the interval.
4. For every fifth value corresponding to an interval, use a diagonal mark on existing four tally marks.
5. Count the tally marks against each interval and record it in the adjacent column titled frequency.

Table 10 shows how tally marks are used for recording the frequency of the variable.

<table>
<tr><th>Number of shopping trips</th><th>Tally Marks</th><th>Frequency</th></tr>
<tr><td>3</td><td>||</td><td>2</td></tr>
<tr><td>4</td><td>~~||||~~ ||||</td><td>9</td></tr>
<tr><td>5</td><td>|||</td><td>3</td></tr>
<tr><td>6</td><td>|||</td><td>3</td></tr>
<tr><td>7</td><td>|</td><td>1</td></tr>
<tr><td></td><td>Total</td><td>18</td></tr>
</table>

Table 10: Tally marks and ungrouped frequency distribution

The table shows maximum tally marks against 4. This implies that the variable "number of shopping trips" takes the data value 4 nine times. In other words, the maximum number of respondents made 4 shopping trips for buying clothes.

Grouped frequency distribution

Often a variable can have a large data set in which the values are spread over a large range, i.e. the difference between the maximum and minimum values is large. In many such cases, using tally marks can also be time-consuming. Therefore, a practical approach is to summarise the data into groups with small intervals. This method of classification is called grouped frequency distribution.

> *"Grouped frequency distribution is a classification technique that is adopted for large data sets."*

Let's understand this with a simple example:

David and Fatima have collected data on "the number of clothes bought" by the respondents for their own use in the last three months.

Ahmed	7	Kartik	5
Alice	7	Li Na	12
Asha	6	Nancy	11
Ben	1	Paul	4
Elizabeth	14	Roger	6
George	5	Samir	5
Jodie	7	Stella	8
John	10	Sue	8
Julie	9	Zoya	8

The raw data has a range from 1 to 14.

Number of clothes bought	Tally Marks	Frequency
1 – 3	|	1
4 – 6	~~||||~~ |	6
7 – 9	~~||||~~ ||	7
10 – 12	|||	3
13 – 15	|	1
	Total frequency	18

Table 11: Grouped frequency distribution

REVISION on the go

Let us look at the grouped frequency distribution presented in Table 11. When you divide the range of values in raw data into groups, the grouped frequency table acquires a structure. The specific elements of this table structure are:

- Class: This is the group of data values. In this example, six classes have been created:

  1–3, 4–6, 7–9, 10–12, 13–15

  As a thumb rule, the number of classes should be between 5 and 10. The reason you would not have less than 5 or more than 10 classes is that it becomes difficult to draw conclusions.
- Class interval: The range of data values in a class. In this example, the class interval is 3.
- Class frequency: The number of instances (items) that fall within a class interval. This can be easily determined with the use of tally marks. In this example, class frequency for the class interval 4–6 is 6.
- Class limits: The lowest and highest values in a class. In this example, the class limits for class 1–3 are 1 (lowest value) and 3 (highest value).
- Total frequency: The total number of instances for all observations (sample). In this example, total frequency is 18.

This table gives some meaningful information. For instance, with a quick look, we understand that the maximum number of instances is in the 7–9 group. Broadly, this means that more than one-third of respondents have purchased 7 to 9 items of clothing in the past three months.

It is a good practice that, wherever possible, a frequency distribution table should have equal classes. This is helpful for comparison. However, in cases where there are only a few instances in the upper or lower ends (also called outliers), these class intervals may be combined to avoid too many classes.

OVER TO YOU

Activity 12: Classification and tabulation

Jiao and Waheeda also collected data on the number of games bought in the past month by 40 respondents.

2	6	14	6	7
8	9	7	9	9
2	6	8	2	12
3	2	15	10	8
7	4	20	18	19
4	10	4	3	6
18	17	1	6	4
6	3	6	5	1

Tabulate the data showing tally marks and class frequencies.

Cumulative frequency distribution

We want to know the number of respondents who bought up to 9 items of clothing in the past three months. This can be done by finding the cumulative frequency. The cumulative frequency distribution is calculated by adding the frequecies successively in a separate column.

Number of clothes bought	Class Frequency	Cumulative Frequency
1 – 3	1	1
4 – 6	6	7
7 – 9	7	14
10 – 12	3	17
13 – 15	1	18
Total frequency	18	

Table 12: Cumulative frequency distribution

REVISION on the go

From the cumulative frequency column in Table 12, you can see that 14 out of 18 respondents have bought up to 9 items of clothing in the past three months.

Cumulative frequency distribution is useful for understanding how many items are equal to a specified value (for ungrouped data), and also less than or up to or more than a specified value (for grouped data).

Relative frequency distribution

Using percentage is a good way to present data because unlike absolute values, percentages indicate the proportion. This type of tabulation is called relative frequency distribution.

Number of clothes bought	Class Frequency	Relative Frequency (%)	Cumulative Frequency	Relative Cumulative Frequency (%)
1 – 3	1	5	1	5
4 – 6	6	34	7	39
7 – 9	7	39	14	78
10 – 12	3	17	17	95
13 – 15	1	5	18	100
Total frequency	18	100		

Table 13: Relative frequency distribution

REVISION on the go

From the relative frequency column in Table 13, you can see that 78% of respondents have bought up to 9 items of clothing in the past 3 months.

OVER TO YOU

Activity 10: Frequency distribution table

Jiao and Waheeda have also collected data on the number of games bought in the past month by 40 respondents.

2	6	14	6	7
8	9	7	9	9
2	6	8	2	12
3	2	15	10	8
7	4	20	18	19
4	10	4	3	6
18	17	1	6	4
6	3	6	5	1

Tabulate the data showing cumulative and relative frequency distribution.

NEED TO KNOW

Statistics is a specialised branch of mathematics consisting of methods that are used to systematically collect, organise and analyse data with the aim to get meaningful information that can be used for decision-making.

There are two branches of statistics: descriptive statistics and inferential statistics.

There are important factors to consider before a statistical study is undertaken: Stating the objective clearly, defining the scope of research and specifying the required accuracy of data.

Data is the physical representation of information in a manner suitable for communication, interpretation or processing by human beings or by automatic means.

Data obtained from pre-existing published or unpublished sources is called **secondary data**. Other sources of secondary data include journals, government reports, television news, and articles on the internet.

The original data that is collected for a research is called **primary data**. Sources of primary data in a business context are surveys of customers, employees and suppliers with the aid of questionnaires, focus groups, personal interviews, and direct observations. Questionnaires are widely used for a variety of surveys.

Secondary and primary data can be either quantitative or qualitative. Quantitative data are numeric observations, and qualitative data are non-numeric (or categorical) observations.

Quantitative data is further classified into discrete data and continuous data.

A variable is any characteristics, number, or quantity that can be measured or counted. A variable may also be called a data item.

Smart managers understand the new environment within which a business exists and use the large volume of data generated each day to take decisions.

The subject of a statistical study is called the population. This population can be people, objects or institutions.

Gathering data from a large population is both time-consuming and costly. A practical way is to collect data from a **sample**.

Data collected from a sample is organised and analysed with the application of statistical methods with the aim to draw conclusions about the population.

The key factors that should be considered in making sample choices are: objectives of the study, nature of the population, definition of the population, research design considerations, availability of resources and ethical and legal considerations.

There are three key approaches to sampling: random sampling, mon-random sampling and mixed sampling.

Random sampling techniques include simple random, systematic, stratified random sampling and cluster (or area) sampling.

Non-random sampling techniques include quota, convenience, judgement and snowball sampling.

Sampling error is the consequence of selecting a sample whose measurement offers a result that does not reflect the true picture of the population characteristic being studied.

Sample size should be optimum. It is determined by two factors: cost and required precision.

Measurement scales for categorising different types of variables include nominal level, ordinal level, interval level and ratio level.

Meaningful information and conclusions can be drawn from this data only after it is classified, presented in the form of tables or charts and diagrams and analysed.

Classification of data is the act of arranging items in groups according to their similarity. Simple classification and complex classification are two key ways of classifying data.

A two-way table is the result of complex classification of data in rows and columns that gives information about two inter-related variables.

Frequency distribution is tabular classification of data for large data sets. A frequency distribution table shows the number of times (or frequency) a variable takes each of its data values.

READING LIST

Anderson, D.R.; Sweeney, D.J.; Williams, T.A.; Freeman, J. and Shoesmith, E. (2014). Statistics for Business and Economics. Cengage Learning, 3rd Ed. ISBN 978–1408072233

Burton, G.; Carroll, G. and Wall, S. (2001). Quantitative Methods for Business and Economics. Financial Times/Prentice Hall, 2nd Ed. ISBN 978–0273655701

Silver, M. (1997). Business Statistics. McGraw Hill Education, 2nd Ed. ISBN 978–9780077092252

RESOURCES: BOOKS

Weiss, N.A. (2013). Introductory Statistics. Pearson, 9th Ed. ISBN 978–1292022017

SOURCES: WEBSITES

Australian Bureau of Statistics Website (2017). Statistical Language – Quantitative and Qualitative Data. http://www.abs.gov.au/websitedbs/a3121120.nsf/home/statistical+language

OCED Website (2013). Glossary of Statistical Terms. https://stats.oecd.org/glossary/index.htm

Chapter 4

Statistical Tools and Data Analysis

Introduction

Chapter 3 introduced you to data, which is the core of statistics. You also saw that raw data is just unorganised numbers or attributes. Raw data becomes useful for drawing meaningful conclusions only after it is processed, summarised and presented. This chapter will begin by taking a step towards visual presentation of data in the form of charts and graphs.

The first section in this chapter will focus on analysis of data by using descriptive statistics. We will also learn to determine correlation between variables and use linear regression for forecasts. It is important that you understand classification and tabulation of data and also various charts and graphs before you begin the section on data analysis.

Learning outcome

4 Analyse data using statistical tools and interpret the results. (Weighting 25%).

Assessment criteria

4.1 Construct and interpret appropriate charts and diagrams from tabular data

4.2 Employ a set of descriptive statistics for analysis and interpretation of grouped and ungrouped data

4.3 Determine correlation between two business variables

4.4 Perform linear regression to make business forecasts

Level 4
Introduction to
Quantitative Methods

4.1 Constructing and interpreting charts and diagrams

Research often throws up large amount of data. In such cases, **charts and graphs** are powerful tools for presenting data because they help to condense data into a visual format. Visual presentation is simple to understand. Charts and graphs are powerful instruments for communicating complicated relationships between variables and presenting complex information. They can leave a lasting impact.

> *"One of the major differences between tables and charts is that a table says "here is your data, now go find the answers"... while a good chart says "here is your answer.""*
>
> *Jorge Camoes cited on UK Government Statistical Service website*

The downside of using graphs is that information presented may not be completely accurate and consistent with the original data. For this reason, a conclusion drawn from a chart or graph may be inconclusive.

You should present charts and graphs for easy reading and interpretation. You can follow these rules or guidelines.

1. Title: a chart or graph should have a meaningful, yet brief, title.
2. The axes: Label both, horizontal *x*-axis (also called abscissa) and vertical *y*-axis (ordinate) properly.
3. Scale: Find a suitable scale for both axes.
4. Data values: Plot these points carefully to reflect information correctly.
5. **Legends**: You should provide these where needed and define scales wherever required. However, avoid excessive details within the chart or graph as it reduces its clarity.
6. Data: supplement the chart or graph with data on which it is based.

The components discussed above are essential for creating meaningful charts and graphs. Following these rules brings clarity and reduces ambiguity in the information being presented. You can see the components depicted in Figure 1.

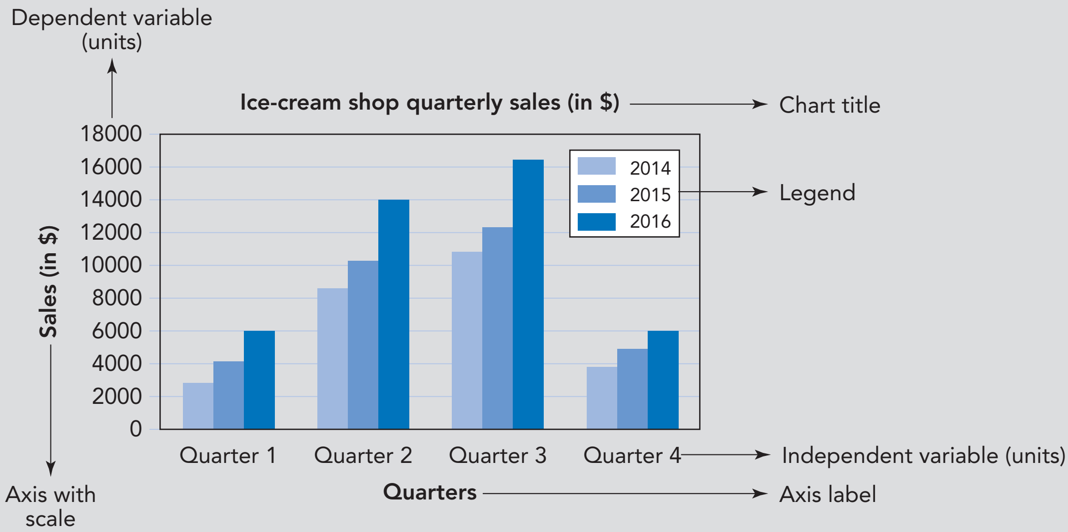


Figure 1: Components of a good chart or graph

There are two broad categories of charts and graphs. These are:

1 Qualitative data charts and graphs
2 Quantitative data charts and graphs

There are various formats of these two types. These are listed in Table 1.

Qualitative data charts and graphs (Constructed from a nominal or ordinal scale of measurement)	Quantitative data charts and graphs (Constructed from an interval or ratio scale of measurement)
• Bar chart	• Histogram
• Pie chart	• Frequency polygon
	• Ogive (Cumulative Frequency Curve)
	• Stem and leaf plot
	• Scatter diagram*

Table 1: Types of charts and graphs

*Scatter diagrams will be discussed in the section on correlation.

Let's understand the other charts and graphs with the help of case study 1.

Charts and graphs for qualitative data

CASE STUDY 1

Chikelu's ice cream shop

Chikelu has a small shop that sells ice cream throughout the year. He sells 6 flavours of ice cream: vanilla, strawberry, butterscotch, chocolate, raspberry and coconut at a fixed price of $2.50 per ice cream. April to September is the summer season in Chikelu's country. Winter months are January to March, and then, October to December.

In January 2017 Chikelu decides to examine his sales data to understand:

1 How sales in 2016 compare on a quarterly basis;

2 How sales in 2016 compare with sales in 2015 and 2014.

Bar chart

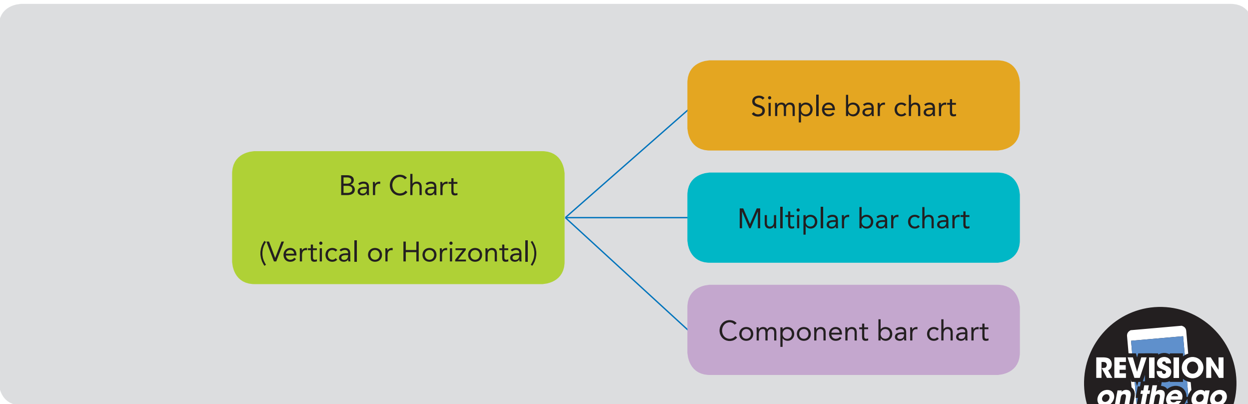
A bar chart is also called a bar graph. Three commonly used formats for a bar chart are:

- Simple bar chart;
- Multiple bar chart;
- Component bar chart.

The properties of a bar chart are:

- It visually presents categorical data and its quantitative values.
- It uses rectangular bars to represent and compare data. There is a degree of separation between each bar or group of bars.
- It is constructed either in vertical (also called column) format or horizontal format.
- It can either be a simple or multiple or component bar chart.
- The categorical data is plotted on the horizontal axis (*x*-axis), and the measurable values are presented on the vertical axis (*y*-axis)

The lengths of the bars are in proportion with the size of the data category represented by them. In other words, the length of each bar determines its value.



Let's construct a simple bar chart for the three years' data Chikelu has collected on his monthly sales of ice cream. The data is presented in Table 2.

	2014 $	2015 $	2016 $
Jan	700	1000	1500
February	900	1450	1800
March	1200	1700	2500
April	2000	2500	3200
May	2900	3700	5000
June	3800	4150	5800
July	3900	4300	6000
August	3900	4200	5700
September	3100	3900	4800
October	1700	2300	2900
November	1200	1700	2000
December	900	900	1200

Table 2: Monthly sales of ice cream

REVISION on the go

Constructing the bar chart

Follow the steps given below to translate Table 2 into a bar chart shown in Figure 2.

1 Add up the monthly sales for 2016 to get quarterly sales, as presented in Table 3.

2 Plot the four quarters on the horizontal axis.

3 Plot the value of sales on the vertical axis on a suitable scale.

4 Carefully construct vertical bars so that each bar corresponds to the correct values.

Quarters	Sales 2016 ($)
Quarter 1 (January–March)	5800
Quarter 2 (April–June)	14000
Quarter 3 (July–Sept)	16500
Quarter 4 (Oct–Dec)	6100

Table 3: Quarterly sales for 2016

REVISION on the go

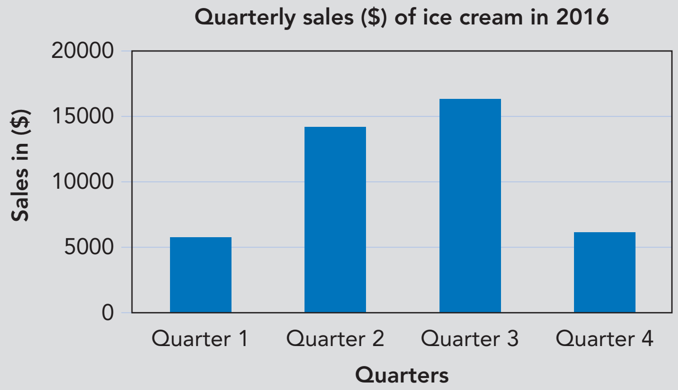


Figure 2: A simple vertical bar chart

Interpreting a simple bar chart

The bar chart in Figure 2 compares sales for four quarters of 2016. You can see clearly here that Chikelu's ice cream shop experienced maximum sales in Quarter 3, which is a period of summer.

When are bar charts appropriate?

Bar charts are good for visual presentation when:

- a comparison needs to be made for sets of data between different groups;
- trends need to be identified.

For example, with a simple bar chart created for the data presented in Table 2, Chikelu can compare the quarterly sales for the four quarters in the year 2016.

Multiple bar chart: vertical format

Chikelu can also compare quarterly sales for the years 2014, 2015 and 2016 with a multiple bar chart. The quarterly sales data for these three years 2014, 2015 and 2016 is presented in Table 4 and depicted in the multiple bar chart in Figure 3.

	Sales for three years ($)		
Years / Quarters	2014	2015	2016
Quarter 1 (January–March)	2800	4150	5800
Quarter 2 (April–June)	8700	10350	14000
Quarter 3 (July–Sept)	10900	12400	16500
Quarter 4 (Oct–Dec)	3800	4900	6100

Table 4: Quarterly sales for three years

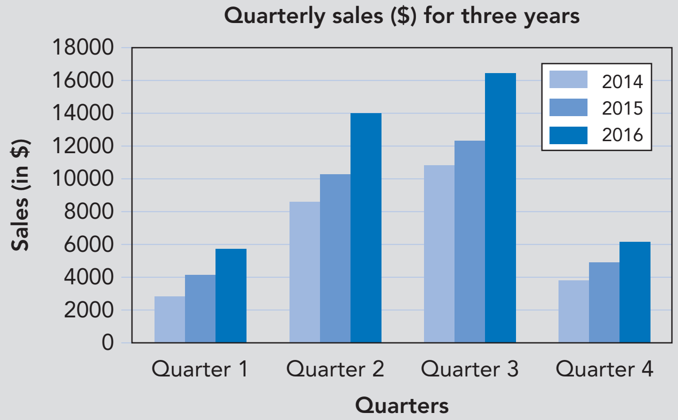


Figure 3: Comparison of quarterly sales for three years

Interpreting the multiple bar chart in Figure 3

The bar chart in Figure 3 compares quarterly sales made by Chikelu over three years. It shows a year-on-year trend. It also shows how the annual sales are distributed throughout the year. Note the use of legend or key that differentiates three different years.

You can see from the bar chart that the ice cream shop experiences highest sales in Quarter 3. This is a consistent pattern in the last three years. The chart also indicates that there is an increasing trend in sales over the last three years in all quarters.

Multiple bar chart: horizontal format

The data in Table 4 can also be presented as a horizontal multiple bar chart as in Figure 4.

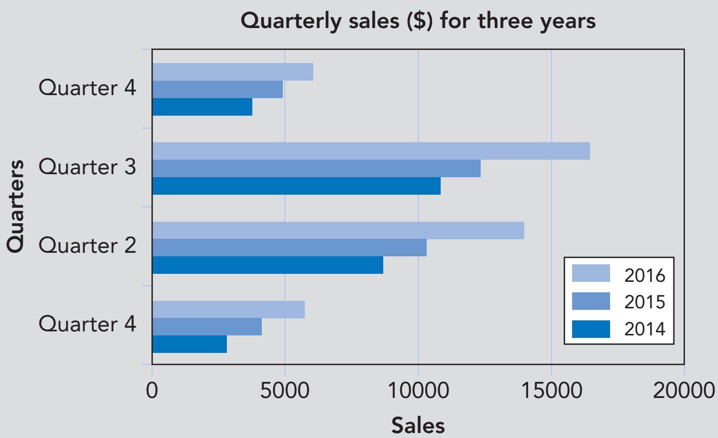


Figure 4: Multiple bar chart in horizontal format

Component bar chart

Table 4 can also be presented as a component bar chart as shown in Figure 5. You can see that in a component bar chart, the four quarters (components of a year) are presented as a single bar. The key elements of this graphical presentation are:

- the legend or key has been added to the graph to show the colours representing different quarters
- the height of each component (defined by the colour) in a bar represents its value or amount.

The components are drawn in the same order across all bars to facilitate comparison.

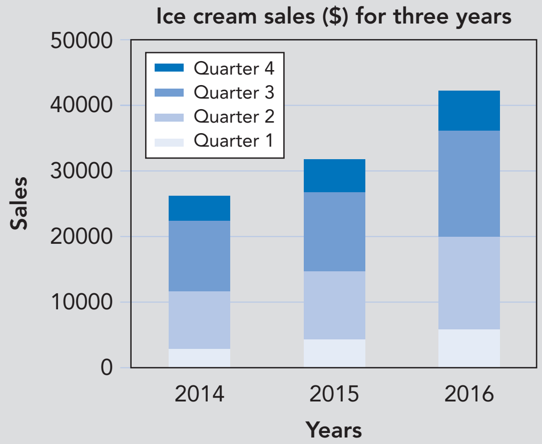


Figure 5: Component bar chart

REVISION on the go

Interpreting the component bar chart in Figure 5

- The height of each bar represents its total value. For example, the top end of 2016 bar is approximately $42,000 on the vertical axis.
- Within each bar, the height of the relevant section defined by its colour represents the component's value. For example, in the year 2016, sales in Quarter 3 were approximately

  $36,000 – $20,000 = $16,000.

This is the difference between the top and bottom end of the coloured segment for Quarter 3.

OVER TO YOU

Activity 1

1 Refer to the data in Table 2 and construct a vertical multiple bar chart to show Chikelu's monthly shop sales in 2014, 2015 and 2016.

2 How would you interpret the bar chart you have drawn?

OVER TO YOU

Activity 2: Component bar chart

Construct a component bar chart to compare the sales made by Chikelu in the summer months (April–September) over the three years 2014, 2015 and 2016. (Extract data from Table 2).

Pie charts

A pie chart is used for presenting categorical data. It also called a pie graph or circular diagram. It resembles a round pizza or pie with slices of various sizes.

A pie chart is an appropriate diagrammatic tool when the numbers of items to be presented are ideally not more than six. Each item is presented as a segment (or a slice), which is a certain proportion of 360°. The slices denote either the value or percentage.

Since a pie chart is circular, it is a presentation of data in 360 degrees (360°). Therefore, in a pie chart: **total of values of all items = 360° = 100%**

CASE STUDY 1 CONTINUED

Chikelu's ice cream shop

Chikelu decides to examine his sales data to understand the quarterly breakup of annual sales for 2016. The data is presented in table 6 below. There are only four items (i.e. four quarters). We have used the data to construct a pie chart with four slices.

Quarters	Sales in 2016 (%)
Quarter 1 (January–March)	5800
Quarter 2 (April–June)	14000
Quarter 3 (July–Sept)	16500
Quarter 4 (Oct–Dec)	6100
Total	42400

Table 5: Quarterly sales for 2016

Follow the steps to create the pie chart for the sales data presented in Table 6.

Step 1: Calculate each quarter's sales as the percentage of total sales.

Quarters	Sales 2016 $	Sales (%)
Quarter 1 (January–March)	5800	$\left(\frac{5800}{42400}\right) \times 100 = 13.67$
Quarter 2 (April–June)	14000	$\left(\frac{14000}{42400}\right) \times 100 = 33.02$
Quarter 3 (July–Sept)	16500	$\left(\frac{16500}{42400}\right) \times 100 = 38.92$
Quarter 4 (Oct–Dec)	6100	$\left(\frac{6100}{42400}\right) \times 100 = 14.39$
Total	42400	100

Table 6: Sales absolute values and percentages

REVISION on the go

Step 2: Convert the calculated percentages into degrees. Each value is a proportion of 360 degrees.

Quarters	Sales 2016 $	Sales (%)	Angle of the slice
Quarter 1 (January–March)	5800	13.67	13.67% × 360 = 49.12 ≈ 49°
Quarter 2 (April–June)	14000	33.02	33.02% × 360 = 118.87 ≈ 119°
Quarter 3 (July–Sept)	16500	38.92	38.92% × 360 = 140.11 ≈ 140°
Quarter 4 (Oct–Dec)	6100	14.39	14.39% × 360 = 51.80 ≈ 52°
Total	42400		

Table 7: Sales absolute values, percentage and angle of slice

REVISION on the go

Step 3: Use the obtained values presented in Table 9 to construct an accurate pie chart.

Quarters	Sales in 2016 (%)	Sales (%)	Segment degrees
Quarter 1 (January–March)	5800	13.67	49°
Quarter 2 (April–June)	14000	33.02	119°
Quarter 3 (July–Sept)	16500	38.92	140°

Quarters	Sales in 2016 (%)	Sales (%)	Segment degrees
Quarter 4 (Oct–Dec)	6100	14.39	52°
Total	42400	100.00	360°

Table 8: Quarterly sales, percentage sales and corresponding degrees for 2016

REVISION on the go

“Angle of the slice in a pie chart = relative frequency of the item × 360°”

Constructing a pie chart manually

Just like any other chart or graph, you can easily construct a pie chart with spreadsheet software. However, for exams, you will need to create it manually. Therefore, to get maximum accuracy, you need to follow these steps:

- Calculate the percentages for each quarter as shown in Table 7.
- Calculate the degree for each segment (quarter) as shown in Table 8.
- Use your compass to draw a circle.
- Use your **protractor** to plot segments of accurate degrees.
- Present the charts with appropriate titles, labels and legend.

The pie chart for absolute values of quarterly sales is shown in Figure 6. The same pie chart can also present the data as percentage sales. This is shown in Figure 7.

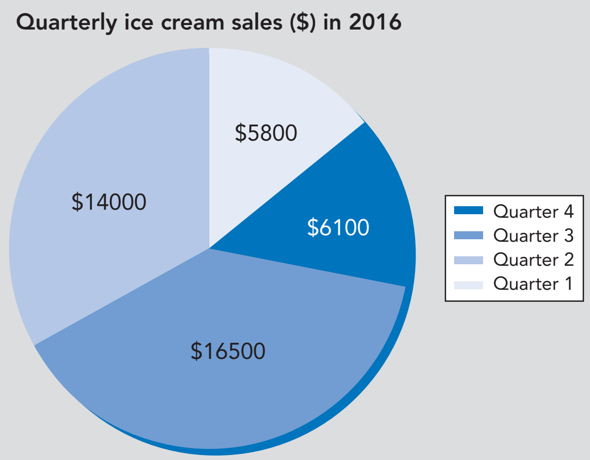


Figure 6: Pie chart presenting sales in absolute numbers

REVISION on the go

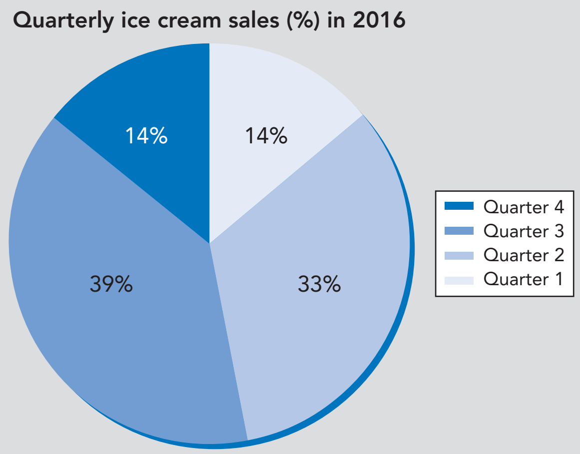


Figure 7: Pie chart presenting sales in percentage

REVISION on the go

Interpreting the pie charts in Figures 6 and 7

The simplicity of a pie chart makes it a powerful tool for presenting information. With a quick glance at the two pie charts in Figures 6 and 7, we understand that both Quarters 2 and 3 have higher sales than Quarters 1 and 4 combined.

OVER TO YOU

Activity 3: Pie charts

Chikelu has his maximum sales in July each year. He wants to break down the $6000 sales for July 2016 by flavour.

	$		$
Vanilla	1100	Chocolate	1800
Strawberry	800	Raspberry	500
Butterscotch	1200	Coconut	600

1 Draw a pie chart to show the share of each flavour.

2 Describe how you would interpret the pie chart you have drawn.

Charts and graphs for quantitative data

Histograms

A histogram displays information in the form of vertical, rectangular bars. The key properties of a histogram ('histo' is a Greek word for area) are:

- It presents data that is numerical, continuous and grouped. The data can be numbers (**absolute frequencies**) or percentages (**relative frequencies**).
- The bars are contiguous (i.e. the bars touch each other). This is because a histogram presents continuous data.
- **Class boundaries** are plotted on the horizontal axis (*x*-axis) and frequency is plotted on the vertical axis (*y*-axis).
- The area of the rectangular bars denotes the frequency for the respective **class intervals**. This gives rise to two possibilities:

1. When class intervals are equal, the width of the bars on the horizontal axis is equal.
2. When the class intervals are unequal, the width and height of the bar for the greater class interval changes, and the dimensions reflect an area that corresponds to the frequency.

Histograms with equal class intervals

Chikelu uses a stopwatch to record the time that customers wait in a queue before being served. The data for 150 customers is presented in Table 9.

Waiting time (seconds)	Number of customers
0–20	5
20–40	15
40–60	16
60–80	45
80–100	40
100–120	15
120–140	8
140–160	6

Table 9: Grouped frequency data

REVISION on the go

"In a histogram, the area (width × height) of each rectangular bar is proportional to the frequency."

Follow the steps given below to translate Table 9 into a histogram that has equal class intervals.

1 Draw the axes and define appropriate scales for both horizontal and vertical axes.
2 Plot the upper class boundaries of the classes (number of ice cream sold) on the horizontal axis.
3 Plot the frequencies (number of days) on the vertical axis.
4 Carefully construct vertical bars so that each bar corresponds to the correct value. Since the class intervals are equal, and height of each bar should correspond to the respective frequency.

The histogram constructed by following the above steps is shown in Figure 8.

Figure 8: A histogram with equal class intervals

REVISION on the go

Histograms with unequal class intervals

Chikelu has tabulated the data on the temperature in February. He wants to draw a histogram from his data.

- The first step is to create groups for the number of ice creams (column A) and record the instances or frequencies (column B) of sale for each group.
- Note the width of each class (column C).
- Decide the standard width as 2 and note the number of standard widths that each class has (column D). Note that a class with a width of 8 has 4 standard widths (2 × 2).
- He then records the frequency density in column E. Frequency density is:

$$\frac{\text{Frequency (Col. B)}}{\text{Number of standard widths (Col. D)}}$$

Temperature (centigrade) (A)	Frequency (B)	Width of each class (C)	Number of standard widths in the class (D)	Frequency density (Col. B ÷ Col. D)
0 to 4	2	4	2	1
4 to 6	10	2	1	10
6 to 10	8	4	2	4

Temperature (centigrade) (A)	Frequency (B)	Width of each class (C)	Number of standard widths in the class (D)	Frequency density (Col. B ÷ Col. D)
10 to 12	2	2	1	2
12 to 14	6	2	1	6

Table 10: Data on temperature in February

- Define the appropriate scale.
- Plot the temperature on the horizontal axis and frequency density on the vertical.

In this example, the standard width is 2, therefore, frequency density implies frequency per two units.

The histogram is depicted in Figure 9.

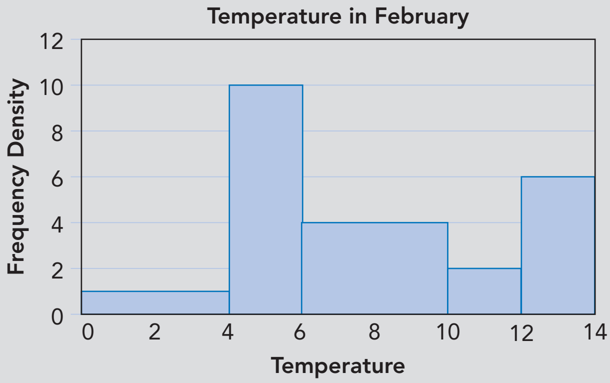


Figure 9: Histogram with unequal class intervals

Histograms

Class intervals are equal

the width of the bars on the horizontal axis is equal. The dimensions reflect an area that corresponds to the frequency.

Class intervals are unequal

the width and height of the bar for the greater class interval changes. The dimensions reflect an area that corresponds to the frequency.

OVER TO YOU

Activity 4: Histogram

Mohammed, a human resource manager, asks 36 workers in the organisation how long it takes them to travel from home to office. He collects the following data:

0 to 10 minutes	5
10 to 20 minutes	12
20 to 30 minutes	7
30 to 40 minutes	6
40 to 50 minutes	4
50 to 60 minutes	2

Construct a histogram using this data.

Frequency polygon

A frequency polygon is extracted from a histogram to show the pattern in frequency distribution of grouped data. A frequency polygon is a line that joins the midpoints of the class intervals plotted on a graph. The data presented in Table 11 is constructed as a frequency polygon in Figure 10.

Frequency polygon is a useful tool for comparing two sets of distributions.

> *A frequency polygon is a line graph constructed by joining the mid-top points (dots) of a histogram, and therefore it is also called the dot-and-line graph. It is a snapshot of the pattern in frequency distribution.*

Waiting time (seconds) (Classes)	Class interval midpoint	Number of customers (Frequency)
0–20	10	5
20–40	30	15
40–60	50	16
60–80	70	45
80–100	90	40
100–120	110	15
120–140	130	8
140–160	150	6

Table 11: Grouped frequency data with class interval midpoints

REVISION on the go

Figure 10: A frequency polygon

Cumulative frequency curve or ogive

As its name suggests, a cumulative frequency curve or ogive (pronounced as o-jive) is used to present the distribution of cumulative frequencies for the given class intervals. This is also a dot-and-line graph in which cumulative frequencies are plotted on the vertical axis and class boundaries are plotted on the horizontal axis. An ogive usually takes the S-shape. Since an ogive starts at the bottom of the frequency (0%) and ends at the sum of all frequencies, the vertical axis of an ogive spans 0% to 100% of the data.

> *An ogive is a graph that presents cumulative frequencies against the upper class boundaries for the classes in a frequency distribution.*

Let's construct an ogive with the data obtained from Chikelu's ice cream shop. The data is presented in Table 12.

Group/Class	Frequency	Cumulative frequency
1–20	5	5
21–40	15	20
41–60	16	36
61–80	45	81
81–100	40	121
101–120	15	136
121–140	8	144
141–160	6	150

Table 12: Daily ice cream sales (April–Sept. 2016)

1 Draw the axes and define the appropriate scales for both horizontal and vertical axes.

2 Plot the upper class boundaries of the classes (number of ice creams sold) on the horizontal axis.

3 Plot the cumulative frequencies (number of days) on the vertical axis.

4 Carefully join the plotted points on the graph by drawing a line that passes through these points. The ogive resulting from Table 12 is shown in Figure 11.

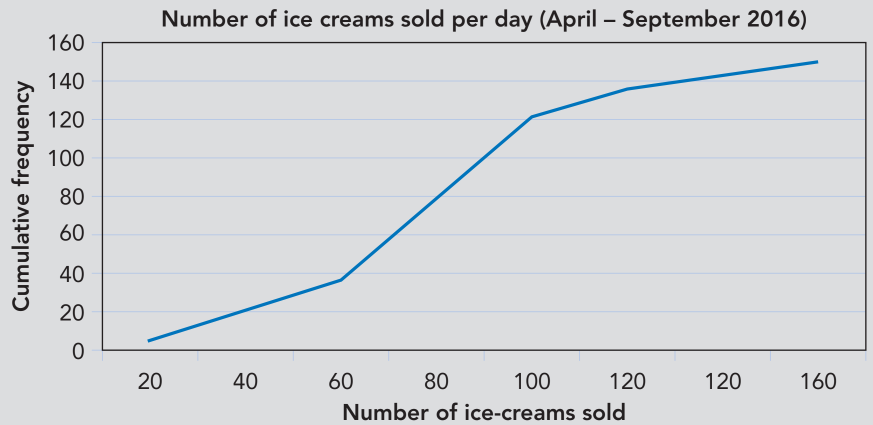


Figure 11: Ogive of daily sales of ice cream.

REVISION on the go

OVER TO YOU

Activity 5: Frequency polygon

Construct a frequency polygon using the data in Activity 4.

Stem and leaf plot

A stem and leaf plot is a diagrammatic summary of raw quantitative data into groups. It is an alternative method to the frequency distribution table. This presentation of data involves breaking up each data value into its leading and trailing digits. The leading digit represents a stem and the trailing digit(s) represent leaf. Let's understand the stem and leaf plot with an example.

In December 2016, Chikelu sold only 480 ice creams in 26 days. His total sales for the month was $1200. The daily unit sales is presented in Table 13.

Unit sales for 26 days in December 2016
11, 7, 12, 8, 17, 8, 25, 17, 15, 13, 24, 15, 6, 22, 9, 28, 33, 12, 15, 22, 31, 42, 45, 21, 13, 9

Table 13:

REVISION on the go

The steps to create a stem and leaf diagrams are:

1 Arrange all stems (leading digits) in a vertical column on the left in the ascending order of magnitude. All items with the same leading digit are grouped within the same stem. For single digit values, the stem is 0.

2 Arrange the leaf component (trailing digit) of each item horizontally in the column adjacent to its corresponding stem. These are also arranged in ascending order.

3 Write down the key to communicate to the reader the meaning of stem | leaf

The stem and leaf diagram for Table 3 is shown in Figure 12.

	Key: 2 | 8 implies 28
0	6, 7, 8, 8, 9, 9
1	1, 2, 2, 3, 3, 5, 5, 5, 7, 7
2	1, 2, 2, 4, 5, 8
3	1, 3
4	2, 5

Figure 12: Stem and leaf diagram

REVISION *on the go*

A stem and leaf plot has many advantages. It clearly communicates:

- Where the maximum frequencies or occurrences lie, i.e. towards upper end or lower end. In this example, the maximum occurrences are in 10 to 20 units.
- The value(s), which occur the most. In Figure 12, three days have recorded sales of 15 units, making it the most frequently occurring units of ice cream sold in a day.
- The range of data, which in this example is from 6 units to 45 units, indicates a high variation in the number of unit sold.
- All data values are used in getting the information. This reduces the loss of information, which typically occurs in frequency distribution, because of using class midpoints for gaining information.

> *A stem and leaf plot is constructed by splitting each data value into two components; the leading digit is called stem and remaining digit(s) are leaf.*

OVER TO YOU

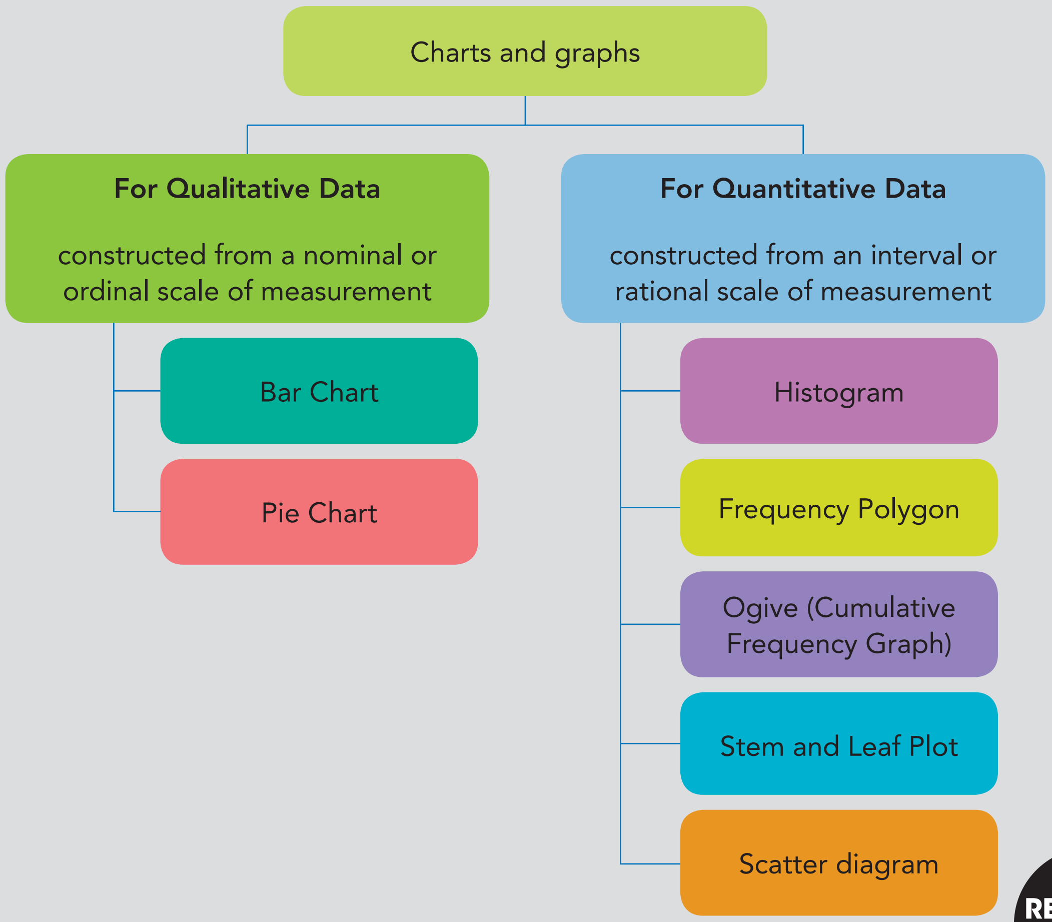
Activity 6: Stem and leaf plot

Theiry is a heavy user of her mobile phone. She realises that her mobile phone bill has increased significantly. She records the call durations of her calls in the past one week. The data in minutes is:

4.6, 8.2, 7.5, 6.8, 4.5, 3.1, 9.3, 4.5, 5.2, 8.6, 6.8, 6.5, 9.7, 2.3, 1.5, 3.8, 5.5, 4.5, 1.8, 3.9

Draw a stem and leaf diagram to show the duration of calls made by Theiry. Interpret the plot.

Classification of charts and graphs



REVISION on the go

4.2 Employing descriptive statistics for analysis and interpretation

Statistical analysis aims to condense unmanageable data into meaningful information. When data is organised through classification and tabulation, and also through diagrams and charts, it helps to draw conclusions about the population or frame. However, to make critical decisions you would need greater insight. This is possible with the application of descriptive statistics.

Descriptive statistics was defined in Chapter 3. As the name suggests, this branch of statistics helps to summarise a data set in a way that it truly represents either the entire population or its sample. Descriptive statistics includes:

- measures of central tendency;
- measures of variability or spread or dispersion.

> *“A measure of central tendency (also referred to as measures of centre or central location) is a summary measure that attempts to describe a whole set of data with a single value that represents the middle or centre of its distribution.”*
>
> *Australian Bureau of Statistics*

> *“Measures of spread describe how similar or varied the set of observed values are for a particular variable (data item). Measures of spread include the range, quartiles and the interquartile range, variance and standard deviation.”*
>
> *Australian Bureau of Statistics*

Let’s understand these measures around these three questions:

1. What are the specific types of measures?
2. Why are we interested in their values?
3. How do we get their values?

Measures of central tendency

In a vast amount of data for a given variable, you should understand what value is representative of the population or frame. For example, a sports club has 100 members of different heights. You could try to find out what is the typical height that represents the club. Is it less than 5.5 ft or between 5.5 ft and 6 ft or more than 6 ft? This information can be found with the measures of central tendency.

There are three measures of central tendency (or location). These are:

1. arithmetic mean (popularly called mean)
2. median
3. mode

There are three instances:

1. when the data is ungrouped;
2. when the data is classified as simple frequency distribution;
3. when the data is classified as grouped frequency distribution.

Arithmetic mean

The arithmetic mean is the average of all values in a data set. Let's understand how to find the arithmetic mean.

Arithmetic mean of ungrouped data

For ungrouped data, the arithmetic mean is calculated simply by dividing the total of the values by the number of observations.

Let's again refer to case study 1 and look at the monthly sales data for Chikelu's ice cream shop. This data is presented in Table 14.

Month	Sales ($) in 2016
January	1500
February	1800
March	2500
April	3200
May	5000
June	5800
July	6000
August	5700
September	4800
October	2900
November	2000
December	1200
Total	42400

Table 14: Ice cream monthly sales

REVISION on the go

This is ungrouped data. Therefore, the arithmetic mean for Chikelu's monthly ice cream sales in 2016 is calculated simply as:

$$\frac{\text{Total annual sales}}{\text{Number of months}}$$

$$\frac{\$42400}{12} = \$3533.33$$

Note that when the mean value of $3533.33 is subtracted from each sales value in Table 14, the sum of the deviations will be 0.

If Chikelu's monthly ice cream sales are denoted as x, his annual sales can be shown as the sum of sales for 12 months. Statistically his total sales for 12 months are represented as $\sum x$ (read as summation x), i.e. $\sum x$ sum of monthly sales in 2016.

If the number of observations is n, then, in this example, n = 12.

Therefore, the formula for calculating arithmetic mean, $\overline{X}$ (also called x-bar), is:

$$\overline{X} = \frac{\sum x}{n}$$

Where,

$\overline{X}$ is the arithmetic mean

$\sum x$ = sum of the values for the item under observation

n = number of observations

In the above example, we have a small discrete data set of n = 12 and this made it easy to determine the arithmetic mean. The formula is also simple. However, in real life, where research generates a large number of observations and mostly relies on frequency distribution (as discussed in Chapter 3), you can determine the arithmetic mean in such cases by using a variation of $\overline{X} = \frac{\sum x}{n}$.

Arithmetic mean of a simple frequency distribution

Mean of simple frequency distribution is determined by the following formula:

$$\overline{X} = \frac{\sum fx}{f}$$

Where,

$\overline{X}$ = mean of x

$\sum fx$ = sum of the product of frequency (f) and x

$\sum f$ = sum of frequencies

Let's calculate the mean using the case study 2.

CASE STUDY 2

At Chikelu's ice cream shop, the maximum sales occur in July. This is a consistent pattern for many years. In July 2016, his shop was open only for 26 days, and the sales amounted to $6000. Chikelu collected daily data on the number of ice creams sold in July 2016.

100, 79, 100, 85, 98, 98, 85, 75, 65, 139, 75, 155, 92, 96, 92, 85, 75, 92, 75, 75, 155, 72, 85, 96, 84, 72

Chikelu wants to find out his mean sales in units on any day when the shop was open in July 2016.

However, because the number of observations is large, i.e. n = 26, there is a greater chance of making an error in calculation of mean. Therefore, it is a good practice to organise the data as frequency distribution.

You can calculate the mean sales by using the formula for arithmetic mean $\overline{X} = \frac{\sum fx}{f}$.

The steps for calculating the mean are:

1 Find the product of x and f for each individual item and write it in the last column.

2 Total the frequency column f and also the product column fx

3 Apply the formula $\overline{X} = \frac{\sum fx}{f}$ on the values calculated in Table 15 to obtain the arithmetic mean.

Units sold (x)	Frequency (f)	Product (fx)
65	1	65
72	2	144
75	5	375
79	1	79
85	5	425
92	3	276
96	2	192
98	2	196
100	2	200
138	1	138
155	2	310
Total	$\sum f = 26$	$\sum fx = 2400$

Table 15

$$\overline{X} = \frac{\sum fx}{f} = \frac{2400}{26} = 92.3 \approx 93$$

The answer is 93 ice creams. The reason why the value is rounded up is because number of ice creams is a discrete data.

Arithmetic mean of a grouped frequency distribution

We have already learnt about grouped frequency distribution. This is the way to classify a large data set into classes and recording the instances (frequency) for each class.

Let us look at Chikelu's ice cream sales data again. Table 16 shows the grouped distribution of the number of ice creams sold (classes) on summer days (frequency) between April and September 2016. Chikelu had kept his shop open for 150 days during the six summer months.

Classes	Frequency
1–20	5
21–40	15
41–60	16

Classes	Frequency
61–80	45
81–100	40
101–120	15
121–140	8
141–160	6

Table 16: Sales for 150 days

REVISION on the go

The method of determining mean in the case of grouped frequency distribution is similar to that for simple frequency distribution, except for one difference. Unlike simple frequency distribution, where an item is expressed as a single value x, in a grouped distribution, x can lie within a range of values or a group organised into class intervals. As a result, we need to assume that all the values in any group are equal to the midpoint of the group. Therefore, we use the midpoint to determine the mean value. This has been illustrated in Table 17.

> *The mean for grouped frequency distribution is determined with the product of mid-point of the classes and the corresponding frequencies.*

The steps are:

- Determine the midpoint of each group or class, taken as x.
- Find the product of x and f for each group and write it in the last column.
- Write the totals of the frequency column f and column with values of fx.
- Apply the formula $\overline{X} = \frac{\sum fx}{f}$ on the total values calculated in Table 17.

Group/Class	Midpoint (x)	Frequency (f)	fx
1–20	10.5	5	52.5
21–40	30.5	15	457.5
41–60	50.5	16	808
61–80	70.5	45	3172.5
81–100	90.5	40	3620
101–120	110.5	15	1657.5
121–140	130.5	8	1044
141–160	150.5	6	903
Total		$\sum f = 150$	$\sum fx = 11715$

Table 17

REVISION on the go

$$\bar{X} = \frac{\sum fx}{f} = \frac{11715}{150} = 78.1 \approx 79$$

An average of 79 ice creams were sold per day between April and September 2016. Note that the final answer 78.1 has been rounded upwards to 79. Since number of ice creams is a discrete data, the answer makes sense.

> *The arithmetic mean for a grouped distribution cannot be determined with 100% certainty. However, the mean can be used for making considerably accurate conclusions in a statistical study.*

Mean of a grouped frequency distribution using assumed mean

Mean of a grouped frequency distribution can also be determined using an assumed mean. This involves the following steps:

Step 1: Estimate the mean of the data and select a value that is close to this estimated value. This is called an assumed mean. This value should be the midpoint of one of the groups.

In the earlier example, the value 90.5 is a good choice to be an assumed mean because it is close to the middle of the distribution. Note that this is also a midpoint of one of the groups/classes.

Step 2: Determine the deviation (d) between x and 90.5, the assumed mean (A). Calculate it for each x and write the value of d in column 4.

The data is presented in Table 18.

Group/Class	Midpoint (x)	Frequency (f)	d = (x – assumed mean)
1–20	10.5	5	–80
21–40	30.5	15	–60
41–60	50.5	16	–40
61–80	70.5	45	–20
81–100	90.5	40	0
101–120	110.5	15	20
121–140	130.5	8	40
141–160	150.5	6	60
Total		$\sum f = 150$	

Table 18

REVISION on the go

Step 3: Take the product of f and d for each group and write the product in the last column.

Group/Class	Midpoint (x)	Frequency (f)	d = (x – assumed mean)	fd
1–20	10.5	5	–80	–400
21–40	30.5	15	–60	–900

Group/Class	Midpoint (x)	Frequency (f)	$d = (x -$ assumed mean$)$	fd
41–60	50.5	16	–40	–640
61–80	70.5	45	–20	–900
81–100	90.5	40	0	0
101–120	110.5	15	20	300
121–140	130.5	8	40	320
141–160	150.5	6	60	360
Total		$\sum f = 150$		$fd = 1860$

Table 19

In this example $\sum fd = -1860$. This is not zero and is also negative. This indicates an overestimation of the assumed mean. This value of mean should be adjusted.

We use the value of $\sum f$ to adjust the assumed mean (A). The formula is $\frac{1}{\sum f} \times \sum fd$

Therefore, the assumed mean of 90.5 is adjusted as:

$$A + \frac{1}{\sum f} \times \sum fd$$

$$90.5 + \frac{1}{150} \times (-1860)$$

$$= 90.5 - 12.4$$

$$= 78.1 \approx 79$$

Note that the mean 79 is same as calculated in the earlier example.

Advantage of arithmetic mean

Arithmetic mean is a simple yet powerful measure of central tendency as it involves all values.

For example, calculating arithmetic mean for monthly sales data implies knowing the value of typical monthly sales. If we know this information, we can use it to compare sales of two stores in the same region. What's more, we can use it to analyse the reasons for which one store has a higher mean sale than the other.

Arithmetic mean can also be used to compare the mean sales for multiple years and analyse the reasons for drop or rise in sales.

Limitation of arithmetic mean

Consider an example. Manju's marks in five subjects in the last term were:

95, 90, 95, 5, 100

Her mean marks will be $\frac{385}{5} = 77$

Note that except for her score of 5 in one subject, she has scored 90 marks or above in all other subjects, and therefore the mean value 77 does not reflect her true performance.

This example illustrates the limitation of arithmetic mean as a measure of central tendency. When the data set is dominated by extremely small or large values (also called outliers), the arithmetic mean is swayed by these outliers and does not reflect the central tendency. The mean in such cases is therefore not representative of the sample or the population.

> *“Arithmetic mean may not be an appropriate measure of central tendency for distributions that have extreme values.”*

> *“Outliers are extreme, or atypical data value(s) that are notably different from the rest of the data.”*
>
> *Australian Bureau of Statistics*

For these distributions, median and mode are better measures for central tendency.

Median

Median is the middle value in a distribution and, therefore, it divides the distribution into two equal halves. In other words, half of the distribution is above the median, and the other half is below it.

Median value is calculated by arranging the data in the order of its magnitude, either in ascending or descending order. The value that emerges in the middle of the arrangement is the median value.

> *“Median is the value that splits the distribution into two equal halves”*

Let look at Manju’s example again. The marks, when arranged in the order of magnitude are:

5, 90, 95, 95, 100

Note that 95 is repeated as a value, yet it is arranged twice, and in the order of magnitude. It is also the value that lies at middle of the arrangement. Therefore, 95 is the median for this data. You can see that the median value (95 marks) is a more accurate measure of Manju’s performance than the mean value (77 marks).

The number of observations can be either odd or even, and therefore the calculation of the middle value of the distribution differs. The formula for each is given in Table 20.

For odd number of observations (n), the median value is	$\frac{n+1}{5}\,th$
For even number of observations, the median value is between	$\frac{n}{2}$ and $\frac{n+2}{2}$

Table 20

Median of a simple frequency distribution

Let us look at Table 21. It has data for the number of ice creams sold by Chikelu in July 2016.

Units sold (x)	Frequency (f)
65	1
72	2
75	5
79	1
85	5
92	3
96	2
98	2
100	2
138	1
155	2
	26

Table 21

REVISION on the go

The total number of observations is 26. Median value will be the middle value, that is, $\frac{26}{2}$ = 13th value. The 13th value can be determined by calculating the cumulative frequencies from Table 22.

Units sold (x)	Frequency (f)	Cumulative frequency
65	1	1
72	2	3
75	5	8
79	1	9
85	5	14
92	3	17
96	2	19
98	2	21
100	2	23
138	1	24
155	2	26
	26	

Table 22

REVISION on the go

From the cumulative frequency values, you can see that the 13th item lies between 9 and 14, or in other words between 79 and 85 units sold.

Therefore, the median value of number of ice creams sold per day in July is 85 as the 13th item will belong to the cumulative frequency 14.

Median of a grouped frequency distribution

The method of determining the median of a grouped frequency distribution is similar to that of simple frequency distribution. The key steps are:

1 calculate the cumulative frequencies;

2 determine the median class.

However, there is one difference. Unlike simple frequency distribution, the exact value of median cannot be obtained. Once the median class is determined, you can calculate the median with the formula:

$$M = l\,\frac{i}{f}\left(\frac{n}{2} - c\right)$$

Where,

l = lower class boundary of median class

i = median class interval

f = frequency of median class

n = total number of observations

c = cumulative frequency of the class preceding median class

Let's apply this method to calculate the median for the number of ice creams sold per day by Chikelu between April and September 2016. The data is presented in Table 23.

Group /Class	Frequency
1–20	5
21–40	15
41–60	16
61–80	45
81–100	40
101–120	15
121–140	8
141–160	6

Table 23: Daily ice cream sales (April – Sept 2016)

Step 1: Find the cumulative frequencies of the grouped distribution as presented in Table 24.

Group	Frequency	Cumulative frequency
1–20	5	5
21–40	15	20

Group	Frequency	Cumulative frequency
41–60	16	36
61–80	45	81
81–100	40	121
101–120	15	136
121–140	8	144
141–160	6	150

Table 24

Step 2: Determine the median class. The total frequency is 150. Therefore, the median is the $\frac{150}{2}$ = 75th observation. From the cumulative frequency column, this observation would lie in the group 61–80 as this group takes observations between 36th and 81st. This is the median class.

Step 3: Finally apply the formula to find the median value.

$$M = l\frac{i}{f}\left(\frac{n}{2} - c\right)$$

$$= 61 + \frac{20}{45}\left(\frac{150}{2} - 36\right)$$

$$= 61 + \frac{20}{45} \times 39$$

$$= 78.33 \approx 79$$

Using an ogive to find the median

You could also determine the median value for the data presented in Table 23 from an ogive or the cumulative frequency graph. The steps are:

1. Translate the data into an ogive or a cumulative frequency graph.
2. Draw a straight line from the 75th point on the vertical y – axis towards the curve.
3. From the point where the straight line touches the curve, draw another straight line towards the horizontal x – axis
4. The median lies at the point where the second straight line meets the x – axis. This is shown in Figure 13.

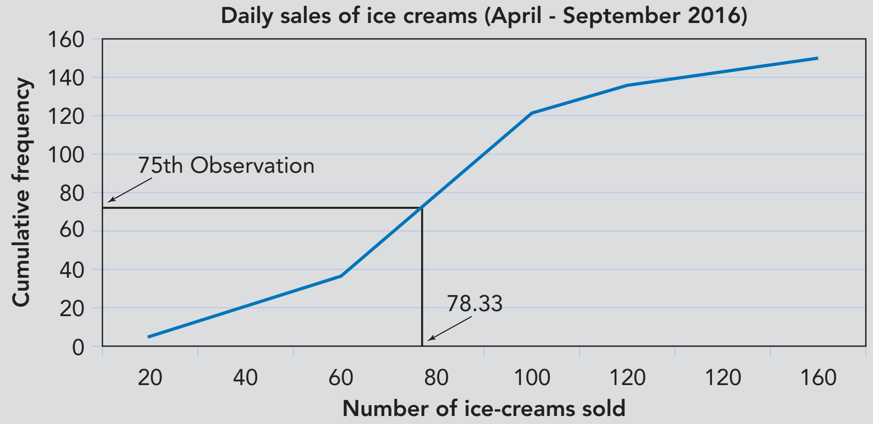


Figure 13: Ogive for number of ice creams sold per day

Mode

It is common for a distribution to have a value that keeps reoccurring. For example, if you ask 100 people about the number of items of clothing they bought in the past one month, perhaps you will get a lot of 2s and 3s in your distribution. It is important to measure the value that occurs most frequently in a distribution. This measure of central tendency is called the **mode**.

> *"The mode is the value that occurs the maximum number of times in a distribution."*

Mode of a simple frequency distribution

Going back to the data on the number of ice creams sold by Chikelu in July 2016 again, let us find the mode value of ice creams sold per day in July. The data is presented in Table 25.

Units sold (x)	Frequency (f)
65	1
72	2
75	5
79	1
85	5
92	3
96	2
98	2
100	2
138	1
155	2
	26

Table 25

Table 25 shows that there are two values that have a frequency of 5, making them the most frequently occurring value in this distribution. Therefore, the mode values for the series are 75 and 85.

As the series has two modes, it is a **bi-modal** series. A series can also be **multi-modal**. The mode is indicated in terms of x and not in terms of the frequencies. Thus, mode is not 5 but 75 and 85.

Mode of grouped frequency distribution

There are two situations:

1 where all groups in the distribution have the same class interval;

2 where all groups in the distribution have the different class interval.

Let's understand how to find the value of mode for these two situations.

Situation 1: Where all groups in the distribution have the same class interval

Let's use this method to find the mode of the number of ice creams sold per day by Chikelu between April and September.

Group/Class	Frequency
1–20	5
21–40	15
41–60	16
61–80	45
81–100	40
101–120	15
121–140	8
141–160	6

Table 26

REVISION on the go

Step 1: Determine the modal class (group) by looking at the frequencies. The group, which has the highest frequency, is the modal class. In this case, the group 61–80 has the highest frequency of 45. Therefore, this is the modal class.

Step 2: Apply the following formula to determine the value of mode.

$$M_o = l + \frac{f_a - f_{a-1}}{(f_a - f_{a-1}) + (f_a - f_{a-1})} \times i_a$$

Where,

l = lower class boundary of modal class

i_a = modal class interval

f_a = frequency of modal class

f_{a-1} = frequency of class immediately preceding the modal class

f_{a+1} = frequency of class immediately following the modal class

Given this,

$$M_o = 61 + \frac{45 - 16}{(45 - 16) + (45 - 40)} \times 20$$

$$= 61 + \frac{29}{29 + 5} \times 20$$

$$= 61 + \frac{29}{34} \times 20 = 78.05$$

Therefore, the mode of number of ice creams sold each day is 78.

Calculating mode where all groups in the distribution have the same class interval.

Step 1: identify the group that has the highest frequencies. It is the modal class.

Step 2: Apply the following formula to determine the value of mode.

$$M_o = l + \frac{f_a - f_{a-1}}{(f_a - f_{a-1}) + (f_a - f_{a-1})} \times i_a$$

l = lower class boundary of modal i_a = modal class interval

f_a = frequency of modal class f_{a-1} = frequency of class immediately before modal class

f_{a+1} = frequency of class immediately following the modal class

Situation 2: Where all groups in the distribution have the different class interval

Chikelu has the following data presented in Table 27 on the number of ice creams bought by his 35 loyal customers over 15 days. He wants to find the mode for this data.

Group	Frequency
0–5	5
5–7	6
7–10	12
10–12	7
13–14	5

Table 27

In the example presented in Table 27, the class intervals are not equal. Mode in such cases can be determined as follows:

Step 1: Find the height (h) of each class by using the formula $h_a = \frac{f_a}{i_a}$

Group	Frequency	Height
0–5	5	1
5–7	6	3
7–10	12	4
10–12	7	3.5
13–15	5	2.5

Table 28

The modal class is the one with the maximum height. In this case, it is the 7–10 group.

Step 2: Apply the following formula to determine the value of mode.

$$M_o = l + \frac{h_a - h_{a-1}}{(h_a - h_{a-1}) + (h_a - h_{a+1})} \times i_a$$

Where,

i = lower class boundary of modal

i_a = modal class interval

h_a = frequency of modal class

h_{a-1} = frequency of class immediately preceding the modal class

h_{a+1} = frequency of class immediately following the modal class

$$M_o = 7 + \frac{4 - 3}{(4 - 3) + (4 - 3.5)} \times 3$$

$$= 7 + \frac{1}{1.5} \times 3$$

$$= 7 + 2$$

$$= 9$$

Therefore, the mode of ice creams bought by Chikelu's 35 loyal customers over 15 days is 9.

Step 1: Find the height (h) of each class.

$$h_a = \frac{f_a}{i_a}$$

The modal class is the one with greatest height.

Step 2: Apply the following formula to determine the value of mode.

$$M_o = l + \frac{h_a - h_{a-1}}{(h_a - h_{a-1}) + (h_a - h_{a+1})} \times i_a$$

Where,

h_a = frequency of modal class

h_{a-1} = frequency of class immediately preceding the modal class

h_{a+1} = frequency of class immediately following the modal class

OVER TO YOU

Activity 7: Mean, median mode

Mohammed, a human resource manager, asks 36 workers in the organisation how long it takes them to travel from home to office. He collects the following data:

0 to 10 minutes	**5**
10 to 20 minutes	**12**
20 to 30 minutes	**7**
30 to 40 minutes	**6**
40 to 50 minutes	**4**
50 to 60 minutes	**2**

Calculate the mean, median and mode for this data.

Measures of dispersion

Consider these two equal sets of data with 10 items each:

Data set 1:	8, 9, 7, 8, 9, 8, 9, 8, 7, 7
Data set 2:	2, 14, 6, 7, 6, 9, 16, 10, 3, 7

Note that:

- The arithmetic mean (central value) for both sets of data is 8.
- However, the spread of data in set 1 is much smaller (items are close to each other and to the central value) than the spread in set 2 (items vary from 2 to 16).

You can see this difference in distribution in Figure 14.

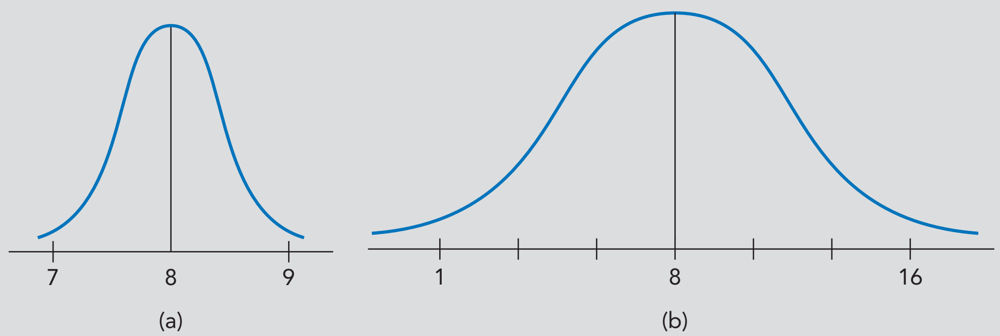


Figure 14: Difference in distribution

REVISION on the go

In a statistical study, it is not only important to have the knowledge of the central location but also of the spread. Therefore, both must be measured.

The measure of spread is also called the measure of dispersion or variability.

There are various measures of dispersion. The focus of this syllabus is on three measures. These are:

1. range
2. inter-quartile range and quartile deviation
3. standard deviation.

Range

Range is the simplest method of measuring the spread of a distribution. It is expressed as the difference between the largest value and the smallest value.

> *Range = Largest value – Smallest value*

Let's consider this example and understand how to determine the value of range.

Chikelu's sister, Naomi, runs a stationery shop. Table 29 presents the quarterly sales for Naomi's shop and Chikelu's shop for the year 2016. Let us compare the data for both shops and determine the range of sales.

	Chikelu $	Naomi $
Quarter 1	5800	7500
Quarter 2	14000	7200
Quarter 3	16500	8000
Quarter 4	6100	8500
Total	42400	31200

Table 29

REVISION on the go

Range is the difference between the highest and the lowest values. Therefore:

	Difference (highest sales – lowest sales)	Value of range
Chikelu's sales	$16500 – $5800	$10700
Naomi's sales	$8500 – $7200	$1300

- Range for Chikelu's sales = $16,500 – $5800 = $10,700
- Range for Naomi's sales = $8500 – $7200 = $1300

You can see that the range is much less for Naomi's sales when compared with Chikelu's sales.

Interpretation: You can see from these values of range that Naomi, although she has lower total annual sales than Chikelu's, has more consistent sales on a month-to-month basis.

The range as a measure of spread has both advantages and disadvantages. Look at Table 30.

Advantages	Disadvantages
Quick to calculate	Heavily influenced by two outliers in widely spread data, and assumes that all other values are evenly distributed
Easy to understand	Ignores variability among other in-between values, and is therefore a weak measure

Table 30

REVISION on the go

You would not often use the value of range for drawing conclusions in business situations.

Inter-quartile range and quartile deviation

This method of measuring dispersion of data is also called semi-quartile range. Inter-quartile range and quartile deviation are based on quartiles Q_1 and Q_3. These measure the deviation between upper and lower quartiles. Quartiles divide the distribution into four parts:

First Quartile Q_1	Second Quartile Q_2	Third Quartile Q_3
Called the lower quartile	Called the median	Called the upper quartile
Has 25% of the distribution, with 1/4th of the distribution below it	Has 50% of the distribution	Has 75% of the distribution, with 3/4th of the distribution below it.

Inter-quartile range is the difference between Q_3 and Q_1.

Quartile deviation can be measured as:

Quartile deviation (QD) = $\frac{Q_3 - Q_1}{2}$

Let's determine the inter-quartile range and quartile deviation of number of ice creams sold by Chikelu each day between April and September 2016. The data is presented in Table 31.

Group	Frequency	Cumulative frequency
1–20	5	5
21–40	15	20
41–60	16	36
61–80	45	81
81–100	40	121
101–120	15	136
121–140	8	144
141–160	6	150

Table 31

REVISION on the go

Step 1: Identify the groups where Q_1 and Q_3 will lie.

Q_1 has 1/4th observations and Q_3 has 3/4th observations in the distribution below them. Since total frequency of the distribution is 150, 1/4th and 3/4th, observations could be calculated as:

$Q_1 = \frac{150}{4}$ = 37.5th observation. This will lie in the group 61–80

$Q_3 = \frac{3(150)}{4}$ = 112.5th observation. This will lie in the group 81–100

Step 2: Calculate the values of Q_1 and Q_3 by using the following formulae.

$$Q_1 = l + \frac{i}{f}\left(\frac{n}{4} - c\right)$$

$$Q_3 = l + \frac{i}{f}\left(\frac{3n}{4} - c\right)$$

Where,

l = lower class boundary of quartile class

i = class interval

f = frequency of quartile class

n = total number of observations

c = cumulative frequency of the class preceding quartile class

Therefore,

$$Q_1 = 61 + \frac{20}{45}\left(\frac{150}{5} - 36\right)$$

= 61 + 0.66

= 61.66

$$Q_3 = 81 + \frac{20}{40}\left(\frac{3(150)}{5} - 81\right)$$

= 81 + 15.75

= 96.75

Like the median, lower and upper quartiles can also be determined from an ogive. This is shown in Figure 15.

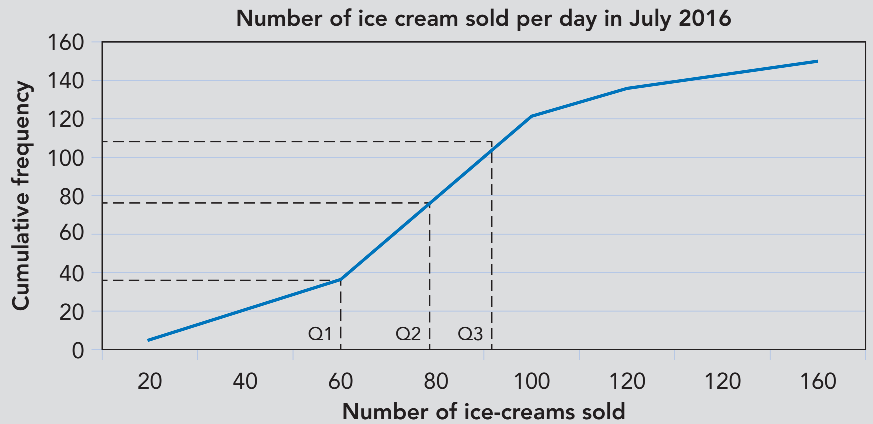


Figure 15: Ogive showing median, lower and upper quartiles

You should note that quartiles derived from a graph are almost never 100% accurate.

Step 3: Apply the formula of inter-quartile range and quartile deviation.

Inter-quartile range = 96.75 – 61.66 = 35.09

Quartile deviation (QD) = $\frac{96.75 - 61.66}{2} = 17.55$

Inter-quartile range and quartile deviation provide a clearer picture of the distribution than range, as these measures remove the extreme values when determining the dispersion. It is generally good practice to use quartile deviation as a measure of spread in cases where central tendency is appropriately reflected by the median rather than the mean.

Standard deviation

This is the most popular measure of dispersion. Simply, deviation implies the distance between a value in a distribution and arithmetic mean of the distribution.
Therefore, if:

X is the value of the observation, and

$\overline{X}$ is the arithmetic mean

Then deviation is $X - \overline{X}$

> *Higher deviation indicates greater spread of data.*

Standard deviation is denoted by the symbol σ (called sigma).

Just like all other measures, standard deviation (σ) can be calculated for ungrouped data, simple frequency distribution and grouped frequency distribution.

The three formulae for calculating standard deviation are shown in Table 32.

Formula for ungrouped data	Formula for simple and grouped frequency distribution	Alternative formula for simple and grouped frequency distribution
$\sigma = \sqrt{\frac{\Sigma (X - \overline{X})^2}{n}}$	$\sigma = \sqrt{\frac{\Sigma f(X - \overline{X})^2}{\Sigma f}}$	$\sigma = \sqrt{\frac{1}{n}\left((\Sigma fx^2) - \frac{(\Sigma fx)^2}{n}\right)}$

Table 32

REVISION on the go

Where,

x is value of the observation

$\overline{X}$ is mean value

n is total observations

σ is standard deviation

If you remove the square root in formula 1, then the term

$\frac{\Sigma (X - \overline{X})^2}{n}$ is referred to as **variance**.

Standard deviation for simple frequency distribution

Let's look at the data on the number of ice creams sold by Chikelu in July 2016 and calculate the standard deviation of the distribution. You can see the data in Table 33.

Units sold (x)	Frequency (f)
65	1
72	2
75	5
79	1
85	5
92	3
96	2
98	2
100	2
138	1
155	2
	26

Table 33

Standard deviation is calculated by using the formula $\sigma = \sqrt{\frac{\sum f(X - \overline{X})^2}{\sum f}}$

For this we will first need the summation of the following values:

Mean $\overline{X}$, deviation from the mean $X - \overline{X}$, square of deviation from the mean $(X - \overline{X})^2$ and product of frequency and square of deviation of the mean $f(X - \overline{X})^2$. The calculations are presented in Table 34.

Units sold (x) Col.1	Frequency (f) Col. 2	fx Col. 3 = Col.1 x Col. 2	$(X - \overline{X})$ Col. 4 = Col.1 – Mean	$(X - \overline{X})^2$ Col. 5 = (Col. 4)^2	$f(X - \overline{X})^2$ Col 6 = Col. 2 x Col. 5
65	1	65	–27.3	745.29	745.29
72	2	144	–20.3	412.09	824.18
75	5	375	–17.3	299.29	1496.45
79	1	79	–13.3	176.89	176.89
85	5	425	–7.3	53.29	266.45
92	3	276	–0.3	0.09	0.27
96	2	192	3.7	13.69	27.38
98	2	196	5.7	32.49	64.98
100	2	200	7.7	59.29	118.58

Units sold (x) Col.1	Frequency (f) Col. 2	fx Col. 3 = Col.1 x Col. 2	$(X - \overline{X})$ Col. 4 = Col.1 – Mean	$(X - \overline{X})^2$ Col. 5 = (Col. 4)^2	$f(X - \overline{X})^2$ Col 6 = Col. 2 x Col. 5
138	1	138	45.7	2088.49	2088.49
155	2	310	62.7	3931.29	7862.58
Total	26	2400		7812.19	13671.54
Mean		92.3			

Table 34

The totals calculated in Table 34 are now inserted in the formula for standard deviation.

$$\sigma = \sqrt{\frac{\Sigma f(X - \overline{X})^2}{\Sigma f}}$$

$$= \sqrt{\frac{13,671.54}{26}}$$

$$= \sqrt{525.83}$$

$$= 22.93$$

Standard deviation for grouped frequency distribution

Let's calculate the standard deviation of the data on number of ice creams Chikelu sold per day between April and September in 2016. The data is presented in Table 35.

Group	Frequency
1–20	5
21–40	15
41–60	16
61–80	45
81–100	40
101–120	15
121–140	8
141–160	6

Table 35

In this case, you could calculate standard deviation using any the formula for standard deviation of grouped distribution.

We will use, $\sigma = \sqrt{\frac{1}{n}\left((\Sigma fx^2) - \frac{(\Sigma fx)^2}{n}\right)}$

For this we will first need the summation of the following values: Mid-values of each group, product of frequency and mid value (fx), square of mid-value (x^2), product of frequency and square of mid value (fx^2).

The data is presented in Table 36.

Group Col. 1	Mid value (x) Col 2 = Col.1 ÷ 2	Frequency (f) Col. 3	fx Col.4 = Col.3 x Col. 2	x^2 Col. 5 = Col 2 x Col. 2	fx^2 Col. 6 = Col. 3 x Col. 5
1–20	10.5	5	52.5	110.25	551.25
21–40	30.5	15	457.5	930.25	13953.75
41–60	50.5	16	808	2550.25	40804
61–80	70.5	45	3172.5	4970.25	223661.3
81–100	90.5	40	3620	8190.25	327610
101–120	110.5	15	1657.5	12210.25	183153.8
121–140	130.5	8	1044	17030.25	136242
141–160	150.5	6	903	22650.25	135901.5
Total		n = 150	11715		1061878

Table 36

Insert the values into the formula for standard deviation.

$$\sigma = \sqrt{\frac{1}{n}(\Sigma fx^2) - \frac{(\Sigma fx)^2}{n}}$$

$$\sigma = \sqrt{\frac{1}{150}\left((1061878) - \frac{(11715)^2}{150}\right)}$$

$$= \sqrt{\frac{1}{150}(1{,}061{,}878 - 914{,}941.5)}$$

$$= \sqrt{\frac{1}{150} \times 146{,}936.5}$$

$$= \sqrt{979.58}$$

$$= 31.30$$

The standard deviation is 31.30

Calculate the deviation for each item

$$X - \overline{X}$$

Compute the square for each deviation

$$(X - \overline{X})^2$$

Calculate variance, i.e. the average of all squared deviations

$$\frac{\sum (X - \overline{X})^2}{n}$$

Calclate standard deviation, i.e. the square root of the variance

$$\sqrt{\frac{\sum (X - \overline{X})^2}{n}}$$

REVISION on the go

Skewness of data

In an ideal distribution, the values are equally distributed on both sides of the mean. The frequency in this case is equal, and when plotted, the distribution shows symmetry. However, in many distributions, a large number of values may gather either on the right side or left side of the mean. This indicates **skewness** of data.

There are three possible distribution patterns you can see in Table 37 and in Figure 16.

Symmetric or not skewed mean = median = mode	The values are placed at equal distance from the mean, and their frequency is equal. The shape of the distribution looks like a bell because mean = median = mode
Right skewed mean > median > mode	The values that are smaller than the median gather closely, i.e. occur with higher frequency than values lower than the median, which have relatively higher distance from the median. This is also called positively skewed distribution.
Left skewed mean < median < mode	The values that are greater than the median gather closely, i.e. occur with higher frequency than values greater than the median, which have relatively higher distance from the median. This is also called negatively skewed distribution.

Table 37

REVISION on the go

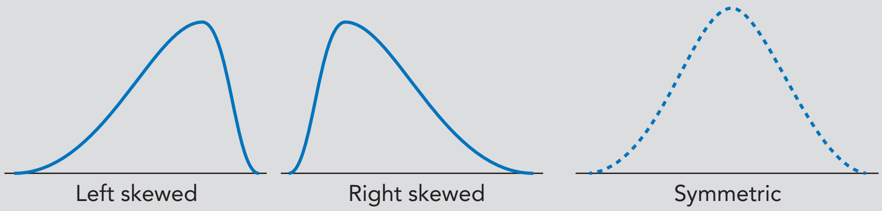


Figure 16: Graphs showing skew

OVER TO YOU

Activity 8

Mohammed, a human resource manager, asks 36 workers in the organisation how long it takes them to travel from home to office. He collects the following data:

0 to 10 minutes	5
10 to 20 minutes	12
20 to 30 minutes	7
30 to 40 minutes	6
40 to 50 minutes	4
50 to 60 minutes	2

Calculate the standard deviation of the data.

Karl Pearson's coefficient of skewness

The skewness of the data centres around the distribution's mean, median and mode. Skewness increases with greater spread of measures of central tendency. Skewness can be measured. One of the methods for measuring the extent of skewness is Karl Pearson's coefficient of skewness.

The formulae for calculating Karl Pearson's coefficient of skewness are:

When values of mean and mode are available	$\text{Pearson's coefficient of skewness} = \dfrac{\text{Mean} - \text{Mode}}{\text{Standard deviation}}$
When values of mean and median are available	$\text{Pearson's coefficient of skewness} = \dfrac{3(\text{Mean} - \text{Median})}{\text{Standard deviation}}$

Table 38

The value of skewness can be as low as –3 or as high as +3. However, value of less than –1 or more than +1 are rare, as these values indicate highly skewed data.

Let's calculate the skewness for the data on number of ice creams sold by Chikelu in July 2016. The data is presented in Table 39.

Units sold (x)	Frequency (f)
65	1
72	2
75	5
79	1
85	5
92	3
96	2
98	2
100	2
138	1
155	2
	26

Table 39

REVISION on the go

The mean, median, mode and standard deviation for this data have been calculated earlier. The values are:

Mean $\overline{X}$ = 92.3

Median = 85

Mode = 75 or 85

Standard deviation = 22.93

Pearson's first coefficient of skewness is:

Pearson's coefficient of skewness = $\frac{92.3 - 85}{22.93} = 0.31$

or

Pearson's coefficient of skewness = $\frac{92.3 - 75}{22.93} = 0.75$

As we have a bi-modal series, Pearson's second coefficient of skewness will be more reliable.

Therefore,

Pearson's coefficient of skewness = $\frac{3(92.3 - 85)}{22.93} = 0.96$

Therefore, we can conclude that data is positively skewed.

4.3 Correlation between variables

In many business problems, it becomes important to establish a **correlation** between two variables. One example of this may be the correlation between age and preference for ice cream.

Correlation is measured by examining the linear relationship between paired data. The value indicates the extent to which variables move together, whether in the same or opposite directions. If variables are correlated, the correlation between them may be either positive or negative. A third situation may be that there is no correlation between two variables at all.

Positive correlation	Two variables move in the same direction. If one increases, the other also increases.
Negative correlation	Two variables move in opposite directions. If one increases, the other decreases.
No correlation	Two variables do not move in any particular pattern and increase or decrease in one is not dependent on the other.

Table 40

The three possible forms of relationship between two variables, x and y, are depicted in the scatter graphs in Figure 17.

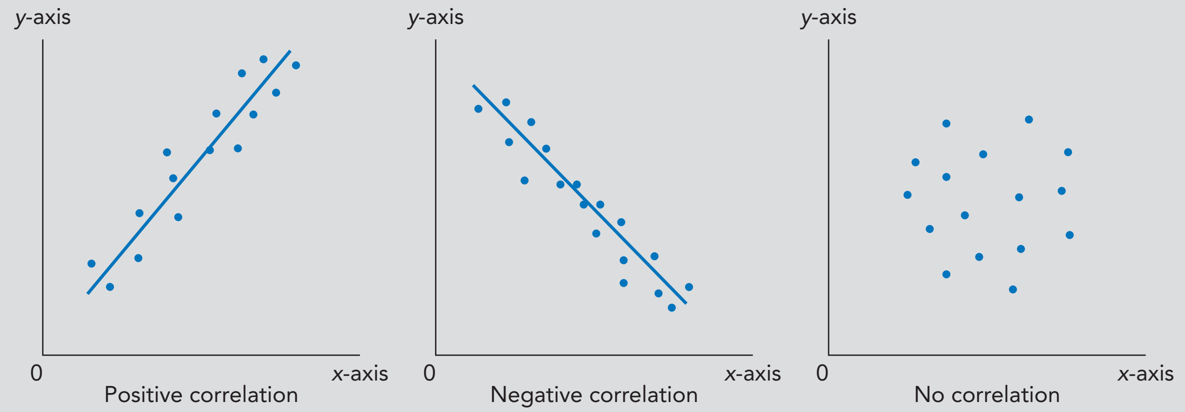


Figure 17: Graphs showing correlation

Scatter graphs to find correlation

A scatter graph or scatter diagrams is a diagrammatic presentation of the relationship between two variables, an independent variable x and a dependent variable y. This shows how a change in the independent variable affects the dependent variable. In other words, it indicates correlation between them.

You would plot the observations for the independent variable on the *x*-axis and observations for the dependent variable on the *y*-axis. The various points plotted on the graph show the kind of relationship between the two variables.

Let's understand scatter graphs (diagrams) with case study 3.

CASE STUDY 3

Fen works in a factory that employs 10 production workers. He is responsible for ensuring the quality of finished products. His job involves carrying out inspections on the finished products and rejecting all units that do not meet the quality standards set by the factory.

In last few months, there has been an increase in the number of units rejected by him. Fen is preparing his monthly report for his employer to explain the reason for the increase in the number of rejections. Fen thinks that there is a positive correlation between a worker's experience and quality. The more experienced workers have fewer rejects.

Fen wants to arrive at a final conclusion, but only after analysing the data. He wants to clearly establish if the number of rejects is dependent on the number of months of experience that a worker has. He also wants to conclude the kind of relationship between the two variables.

Fen's collects one week's data for 10 factory workers. You can see this data in Table 41.

Factory worker	Experience (in months)	Rejects per 100 units
1	12	22
2	11	23
3	16	18
4	3	35
5	10	25
6	14	21
7	6	36
8	13	21
9	9	28
10	8	31

Table 41

REVISION on the go

Let's look at the given data to analyse whether the number of rejects (dependent variable y) is affected by the experience of a worker (independent variable x). For this, let's draw a scatter graph by following the given steps.

1 Plot the independent variable (experience of workers) for each worker on the x-axis.

2 Plot the dependent variable (number of rejects) for each worker on the y-axis.

3 Draw a line of best fit. This line passes almost through the middle of the plotted points.

Note: You should draw the line in such a way that most of the points lie on the line, and the number of points above the line are almost the same as the number of points below it. The strength of correlation depends on how close the points are to the line of best fit.

The closer the points are to the line, the stronger is the correlation between the two variables.

You can see the scatter graph (diagram) showing the relationship between experience of a worker and the number of rejects in Figure 18.

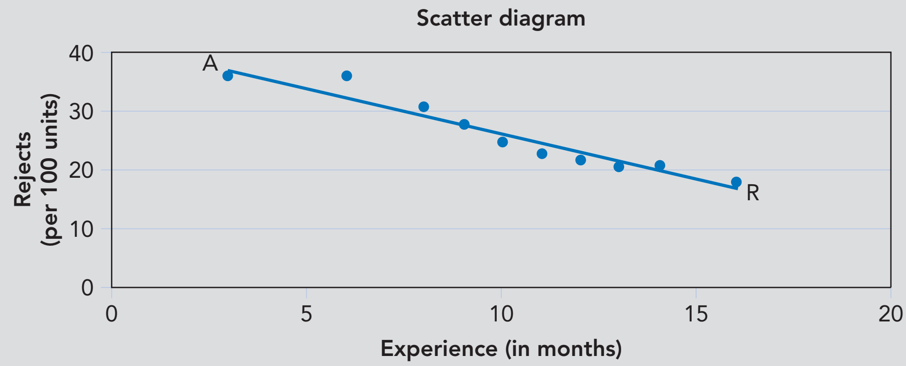


Figure 18: Scatter diagram

Interpreting the scatter graph (diagram)

The scatter graph (also called scatter plot) in Figure 18 shows that observations are close to the line of best fit (labelled as line AB). It clearly indicates that:

- there is a definite relationship between the two variables;
- there is a linear relationship;
- there is negative correlation because the line of best fit is sloping downwards.

This implies that workers with greater experience have lesser numbers of rejects.

Measurement and interpretation of correlation

Pearson's coefficient of correlation

Pearson's coefficient of correlation is more commonly known as **coefficient of correlation** or **correlation coefficient**. It is a statistical measure that gauges the association between paired data for two variables. It quantifies the direction and strength of the linear relationship.

> *"Pearson's coefficient of correlation measures the linear relationship between variables. When there is a change in one variable, it results in a proportional change in the other variable."*

It is represented by the letter r or R. The value of r or R lies between −1 and +1.

- A positive value of the coefficient indicates a positive relationship between two variables or paired data. The closer it is to +1 the stronger is the positive correlation. For example, a correlation coefficient of $r = 0.90$ suggests a strong, positive linear association between two variables.

A coefficient of +1 implies perfect positive correlation.

- A negative value of the coefficient indicates a negative relationship between two variables or paired data. For example, a correlation of $r = -0.3$ suggests a weak, negative association between two variables or paired data. The closer it is to −1 the stronger is the negative correlation.

  A coefficient of –1 implies perfect negative correlation.

- If the coefficient is equal to 0, it indicates that there is no linear relationship or correlation.

You can see this in Figure 19.

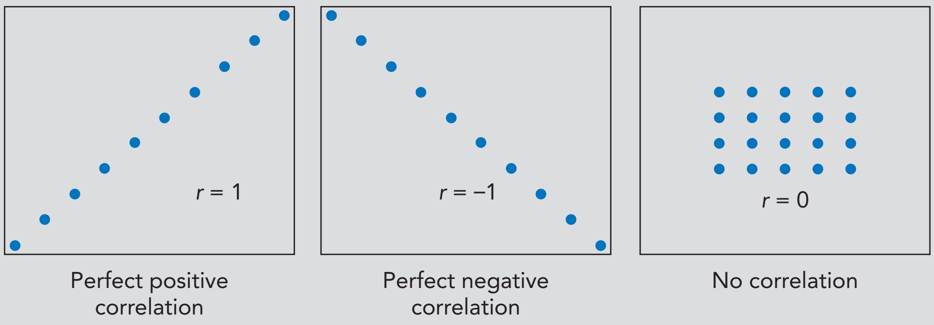


Figure 19: Graphs showing correlation coefficient

Pearson's coefficient of correlation can be measured by using the following formula:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Where,

x = Observations on independent variable

y = Observations on dependent variable

n = Number of observations

Let's refer to case study 3 and calculate the value of Pearson's coefficient of correlation for data presented in Table 42. We will need the following values for calculating the coefficient.

$\sum x, \sum y, \sum xy, \sum x^2, \sum y^2$

These have been calculated in Table 42.

Production worker (Col. 1)	Experience (x) (Col.2)	Rejects (y) (Col. 3)	xy (Col.2 x Col. 3)	x^2 (Col. 2 x Col. 2)	y^2 (Col.3 x Col.3)
1	12	22	264	144	484
2	11	23	253	121	529
3	16	18	288	256	324

Production worker (Col. 1)	**Experience (x) (Col.2)**	**Rejects (y) (Col. 3)**	**xy (Col.2 x Col. 3)**	**x^2 (Col. 2 x Col. 2)**	**y^2 (Col.3 x Col.3)**
4	3	35	105	9	1225
5	10	25	250	100	625
6	14	21	294	196	441
7	6	36	216	36	1296
8	13	21	273	169	441
9	9	28	252	81	784
10	8	31	248	64	961
Total	$\sum x = 102$	$\sum y = 260$	$\sum xy = 2443$	$\sum x^2 = 1176$	$\sum y^2 = 7110$

Table 42

$$r = \frac{10 \times 2443 - (102)(260)}{\sqrt{[10 \times 1176 - (102)^2][10 \times 7110 - (260)^2]}}$$

$$= \frac{24430 - 26520}{\sqrt{[11{,}760 - 10{,}404][71100 - 67{,}600]}}$$

$$= \frac{-2090}{\sqrt{1356 \times 3500}}$$

$$= \frac{-2090}{\sqrt{4{,}746{,}000}} = -0.95$$

A correlation coefficient of $r = -0.95$ suggests a strong, negative linear association between experience of workers and number of rejects of goods manufactured by them.

Spearman's rank coefficient of correlation

Spearman's rank coefficient of correlation or Spearman's correlation coefficient, represented by symbol r or ρ (*rho*), measures the strength and direction of association between the rankings of two variables. This method is used to measure the association between variables when the variables tend to change together, but the rate of change may not necessarily be constant. This method first ranks the raw data and then measures the correlation of ranked values. Like Pearson, Spearman correlation coefficients can range in value from −1 to +1.

Rank correlation is measured by using the following formula:

$$r = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

Where

d = difference between ranking of the two data sets

n = number of paired data or observations

CASE STUDY 4

Farah has data on marks scored by her students in Financial Management and Quantitative Methods. She thinks that only those candidates who perform well in Quantitative Methods get high marks in Financial Management. In other words, she thinks that there is a correlation.

She takes a sample of 12 students and decides to determine the relationship, if any, between the performances of candidates in the two subjects. The data is presented in Table 43.

Student	Quantitative Methods (Max. marks 100)	Financial Management (Max. marks 100)
1	50	45
2	85	37
3	67	86
4	87	78
5	63	58
6	74	60
7	25	48
8	57	57
9	80	68
10	59	92
11	98	88
12	79	70

Table 43

Let's calculate Spearman's rank correlation for the data in Table 43. The steps are:

1 Create a table with 7 columns. The first two columns contain the data whose correlation needs to be measured.

2 Rank the marks scored by candidates in quantitative methods. Enter the ranks in R1 column. Ranking '1' implies highest marks and lowest marks get the lowest ranking.

3 Rank the students on the same basis on their marks in Financial Management and enter the ranks in R2 column.

4 Find the difference in the ranks (*d*) of each row by subtracting R2 from R1. Square the value of each (*d*) to remove the negative values and enter in the last column.

	Quantitative Methods (Max. marks 100)	Financial Management (Max. marks 100)	Quantitative Methods Rank (R1)	Financial Management Rank (R2)	d = R1 – R2	d^2 = d x d
1	50	45	9	11	–2	4
2	85	37	3	12	–9	81
3	67	86	7	3	4	16
4	87	78	2	4	–2	4
5	63	58	8	8	0	0
6	74	60	6	7	–1	1
7	25	48	10	10	0	0
8	57	57	9	9	0	0
9	80	68	4	6	–2	4
10	59	92	8	1	7	49
11	98	88	1	2	–1	1
12	79	70	5	5	0	0
						$\sum d^2 = 160$

Table 44

Finally, use $r = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$ to determine the Spearman's correlation coefficient.

$$r = 1 - \frac{6 \times 160}{12(12^2 - 1)}$$

$$r = 1 - \frac{960}{1716} = 0.44$$

The value 0.44 indicates a positive correlation between performance of candidates in quantitative methods and financial management.

4.4 Performing linear regression to make business forecasts

In the previous section, we discussed how we could identify the correlation between two variables and measure the strength of their association by calculating the correlation coefficient. Simple **linear regression** goes a step further and attempts to fit a linear equation to the observed data on the two variables, an independent variable x and a dependent variable y for modelling their relationship.

The independent variable is also called **predictor** or **explanatory** variable and the dependent variable is called **response** or **outcome.** The independent variable is used to extrapolate data and make business forecasts.

> *Simple linear regression is a statistical method used for analysing the association between two quantitative variables for business modelling.*

A simple linear regression line is expressed as equation of a straight line $y = a + bx$ (equation of a straight line was discussed in Chapter 2).

In the equation x is the independent variable, y is the dependent variable, b is the slope of the linear regression line and a is the y-**intercept**.

Least-squares regression

The use of the least squares method is the most popular method of fitting a simple regression line. Broadly, the method calculates the line of best fit for the observed data in a way that we can minimise the sum of the squares of the vertical deviations from each data point to the line. Vertical deviations of any point on the line of best fit are zero.

Method 1

We can find the line of best fit for a relationship by finding the values of a and b using the formula

$a = \bar{y} - b\bar{x}$

Where, $b = \dfrac{\sum(x - \bar{x}) \times (y - \bar{y})}{\sum(x - \bar{y})}$

These values can be put in the equation $y = a + bx$ to find the simple linear regression equation.

Let's again refer to case study 4. In the previous section, we identified a strong negative correlation between experience of production workers and the number of rejects. We can now fit a least square regression line to express this relationship so that we can forecast the number of units that are likely to be rejected at specific levels of experience of production workers.

Production worker	Experience (in months) x	Rejects per 100 units y	$x - \bar{x}$	$y - \bar{y}$	$((x - \bar{x}) \times (y - \bar{y}))$	$(x - \bar{x})^2$
1	12	22	1.8	–4	–7.2	3.24
2	11	23	0.8	–3	–2.4	0.64
3	16	18	5.8	–8	–46.4	33.64
4	3	35	–7.2	9	–64.8	51.84
5	10	25	–0.2	–1	0.2	0.04
6	14	21	3.8	–5	–19	14.44
7	6	36	–4.2	10	–42	17.64
8	13	21	2.8	–5	–14	7.84
9	9	28	–1.2	2	–2.4	1.44
10	8	31	–2.2	5	–11	4.84
	$\sum x = 102$	$\sum y = 260$			$\sum(x - \bar{x}) \times (y - \bar{y}) = -209$	$\sum(x - \bar{x})^2 = -135.6$

Table 45

Step 1: Find the mean value of x and mean value of y

$$\bar{x} = \frac{\sum x}{n} = \frac{102}{10} = 10.2$$

$$\bar{y} = \frac{\sum y}{n} = \frac{260}{10} = 26$$

Step 2: Find the slope of the line of best fit, b, for this relationship by using the formula

$$b = \frac{\sum(x - \bar{x}) \times (y - \bar{y})}{\sum(x - \bar{x})^2}$$

For this we need the summation of following values,

$x - \bar{x}$, $y - \bar{y}$, $(x - \bar{x}) \times (y - \bar{y})$ and $(x - \bar{x})^2$

Calculation of these is shown in Table 45.

Therefore, $b = \frac{-129}{135.6} = -1.54$

Step 3: Use $b = -1.54$ in the formula $a = \bar{y} - b\bar{x}$ and find the y-intercept a.

$a = \bar{y} - \mathrm{b}\bar{x}$

$= 26 - (-1.54)10.2$

$= 26 + 15.71 = 41.71$

Step 4: Insert the values of b and a in the equation $y = a + bx$ to get the simple linear regression equation.

Therefore, $y = 41.71 - 1.54x$

Method 2

The alternate method is similar to method 1 as for calculation of y-intercept a. The difference is in calculation of the slope b.

Under this method, the slope b is calculated using the formula

$$b = \frac{(n\sum xy - (\sum x)(\sum y))}{(n\sum x^2 - (\sum x)^2)}$$

Production worker	Experience (in months) x	Rejects per 100 units y	xy	x^2
1	12	22	264	144
2	11	23	253	121
3	16	18	288	256
4	3	35	105	9
5	10	25	250	100
6	14	21	294	196
7	6	36	216	36
8	13	21	273	169

Production worker	Experience (in months) x	Rejects per 100 units y	xy	x^2
9	9	28	252	81
10	8	31	248	64
	$\sum x = 102$	$\sum y = 260$	$\sum xy = 2443$	$\sum x^2 = 1176$

Table 46

Therefore,

$$b = \frac{(10 \times 2443 - (102)(260))}{(10 \times 1176 - (102)^2)}$$

$$= \frac{(24{,}430 - 26{,}520)}{(11{,}760 - 10{,}404)}$$

$$= \frac{-2090}{1356} = -1.54$$

$$a = \bar{y} - b\bar{x}$$

$$= 26 - (-1.54)10.2$$

$$= 26 + 15.71 = 41.71$$

Therefore, $y = 10.29 - 1.54x$

Using regression equations

Regression equation can be used in extrapolating of data and forecasting. Businesses use regression equations for correlated variables to estimate the value of one variable corresponding to another variable.

Let's again refer to case study 3. Fen's employee Doreen was temporarily moved to a different department last week. Fen asked her to take on the production role again in the factory. Doreen has 18 months of experience as a factory worker. Fen is moving Doreen in place of a factory worker who has only 3 months of experience, and has an average of 35 rejects per 100 units manufactured. What reduction in rejects can Fen expect from Doreen?

We can use the simple regression equation derived earlier for working out the number of expected rejects from Doreen.

$y = 41.71 - 1.54x$

Doreen's experience is the independent variable = x = 18 months

If we input $x = 18$ in the linear regression equation $y = 41.71 - 1.54x$, we can find the value of y which indicates expected number of rejects.

$$y = 41.71 - 1.54x$$

$$= 41.71 - 1.54 \times 18$$

$= 13.99 \approx 14$ **rejects**

Therefore, expected number of rejects for Doreen once she re-joins the factory is 14 per 100 units manufactured by her. This will be a reduction of 35 − 14 = 21 rejects, and therefore the move is highly recommended.

! NEED TO KNOW

Charts and graphs are powerful tools for presenting data because they help to condense data into a visual format. Visual presentation is simple to understand.

You should present charts and graphs for easy reading and interpretation. There should be a meaningful **title** of the chart or graph. Label the axes properly. The **scale** should be suitably depicted for both the axes and the **data values** plotted carefully to reflect information correctly. Provide legends but avoid excessive details within the chart or graph as it reduces its clarity. A chart or graph should be supplemented with data on which it is based.

There are two broad categories of charts and graphs: qualitative data charts and graphs (bar chart and pie chart) and quantitative data charts and graphs (histogram, frequency polygon, ogive, stem and leaf, scatter).

A bar chart can be a simple bar chart, multiple bar chart or component bar chart.

Bar charts visually present categorical data and its quantitative values. They are constructed either in vertical (also called column) format or horizontal format.

Bar charts are good for visual presentation when a comparison needs to be made for sets of data between different groups or trends need to be identified.

Use a pie chart for presenting categorical data. It also called a pie graph or circular diagram. It is an appropriate diagrammatic tool when the numbers of items to be presented are not more than six.

A histogram shows information in the form of vertical, rectangular bars. It presents data that is numerical, continuous and grouped. The data can be numbers (absolute frequencies) or percentages (relative frequencies). The bars touch each other and area of the rectangular bars denotes the frequency for the respective class intervals.

A frequency polygon is a line graph constructed by joining the mid-top points (dots) of a histogram, and therefore it is also called the dot-and-line graph. It is a snapshot of the pattern in frequency distribution.

An ogive is a graph that presents cumulative frequencies against the upper class boundaries for the classes in a frequency distribution.

A stem and leaf plot is a diagrammatic summarisation of raw quantitative data into groups. It is an alternative method to frequency distribution table.

A stem and leaf plot is constructed by splitting each data value into two components; the leading digit is called stem and remaining digit(s) are leaf.

Descriptive statistics helps to summarise a data set in a way that it truly represents either the entire population or its sample. These include measures of central tendency and measures of variability or spread.

A measure of central tendency is a summary measure that attempts to describe a whole set of data with a single value that represents the middle or centre of its distribution. The three measures of central tendency are arithmetic mean, median and mode.

The arithmetic mean is **average** of all values in a data set. Median is the value that splits the distribution into two equal halves. The mode is the value that occurs the maximum number of times in a distribution.

For ungrouped data, the arithmetic mean is calculated simply by dividing the total of the values by the number of observations.

The mean for grouped frequency distribution is determined with the product of midpoint of the classes and the corresponding frequencies.

Measures of spread describe how similar or varied the set of observed values are for a particular variable (data item). Measures of spread include the range, quartiles and the interquartile range, variance and standard deviation.

Range is the simplest method of measuring the spread of a distribution. It is expressed as the difference between the largest value and the smallest value.

Inter-quartile range and quartile deviation measure the deviation between upper and lower quartiles. These provide a clearer picture of the distribution than range as these measures remove the extreme values when determining the dispersion.

Standard deviation is the most popular measure of dispersion. Higher deviation indicates greater spread of data.

There are three possible distribution patterns:

mean = median = mode, the distribution is not skewed

mean > median > mode, the distribution is right skewed

mean < median < mode, the distribution is left skewed.

Karl Pearson's coefficient of skewness is used for measuring the extent of skewness in a distribution.

Correlation is measured by examining the linear relationship between paired data. If variables are correlated, the correlation between them may be either positive or negative.

Scatter graphs are diagrammatic presentation of the relationship between two variables, an independent variable x and a dependent variable y.

A scatter graph shows how a change in the independent variable affects the dependent variable. In other words, it indicates correlation between them.

Pearson's coefficient of correlation measures the linear relationship between variables. When there is a change in one variable, it results in a proportional change in the other variable.

Correlation is represented by the letter r or R The value of r or R lies between −1 and +1.

Spearman's rank coefficient of correlation or Spearman's correlation coefficient, represented by symbol r or ρ(*rho*) , measures the strength and direction of association between the rankings of two variables.

Simple linear regression is a statistical method used for analysing the association between two quantitative variables for business modelling.

The use of least squares method is the most popular method of fitting a simple regression line.

Businesses use regression equation for correlated variables to estimate the value of one variable corresponding to another variable.

READING LIST

Anderson, D.R.; Sweeney, D.J.; Williams, T.A.; Freeman, J. and Shoesmith, E. (2014). Statistics for Business and Economics. Cengage Learning, 3rd Ed. ISBN 978–1408072233

Burton, G.; Carroll, G. and Wall, S. (2001). Quantitative Methods for Business and Economics. Financial Times/Prentice Hall, 2nd Ed. ISBN 978–0273655701

Silver, M. (1997). Business Statistics. McGraw Hill Education, 2nd Ed. ISBN 978–9780077092252

RESOURCES: BOOKS

Weiss, N.A. (2013). Introductory Statistics. Pearson, 9th Ed. ISBN 978–1292022017

Glossary

Absolute frequency is the number of instances/occurrences observed for a particular value for a variable.

Antecedent is the first term in a given ratio or the first and third terms in a given proportion. For example, in the ratio 2 : 3, the value 2 is the antecedent. Similarly in a proportion, 2 : 3 :: 4 : 5, the values 2 and 4 are antecedents. Also see **consequent**.

Arithmetic operation is a mathematical process performed on numbers. The basic operations are add, subtract, multiply and divide (+, –, x and ÷). Examples of more complex operations include squares, cubes, square root, cube root.

BEDMAS is an acronym for the rule for order of operations in arithmetic and algebra. Also see **order of operations**.

Bias is a deliberate step in the process of statistical analysis that cause results or conclusions which are not a true representation. Bias is different from random sampling error. An example of bias is 'sample selection bias' that systematically chooses non-random data based on particular attributes.

Chart or graph is a visual representation of data in various forms such as rectangular bars, lines or a circular pie. Information is easily and quickly gained with the help of a chart or a graph.

Class boundary the boundary that separates two classes or groups of data. A class has an upper class boundary and a lower class boundary.

Class interval is the range of a class or a group of data.

Coefficient a number that is attached to and multiplies a variable in an algebraic expression. For example, in the equation $ax^2 + bx + c = 0$, x is a variable, while a and b are coefficients.

Compound interest in the first instance of calculation, it is always equal to simple interest. From second instance of calculation, compound interest is the interest earned or paid on the original amount or principal plus any interest accumulated in previous period(s). It is based on a specific rate percentage.

Consequent is the second term in a given ratio or the second and fourth terms in a given proportion. For example, in the ratio 2 : 3, the value 3 is the consequent. Similarly in a proportion, 2 : 3 :: 4 : 5, the values 3 and 5 are consequents. Also see antecedents.

Constant is the quantity that does not change in a mathematical expression. For example, in the equation $x + 7 = y$, 7 is a constant.

Correlation is a statistical approach that measures the strength of association between two variables.

Denominator is the non-zero divisor in a fraction. If $\frac{a}{b}$ is a fraction then b is the denominator.

Dependent variable is a variable that is affected by change in another variable. For example, demand of a product may be affected by the variable price. Also see **response**.

Equation is a specific type of algebraic expression that has arithmetic operation(s) and an equal sign. For example, $2X + 2 = 8$ is an equation. Similarly, $A = 2l + 2w$ is an equation.

Exponent is power used on a base in an exponential notation. For example, in exponential notation 5^3, 3 is the exponent on 5, and the expression indicates 5 is used three times as a factor.

Foreign exchange is the exchange of a country's currency for another country's currency based on current market rates. The system is useful for international trade and travel.

Gradient is the slope of a straight line calculated as change in vertical length divided by change in horizontal length or $\frac{\text{change in } y}{\text{change in } x}$.

Independent variable is a variable that is not affected by change in any other variable. For example, age or gender. Also see **predictor**.

Legend is a brief identification of data presented in a chart, graph or table.

Linear equation is the first order equation consisting of one or more variables with exponent 1. It is generally expressed as $ax + b = 0$. A linear equation when plotted on a graph gives a straight line.

Linear regression is the method of measuring the effect on behaviour of the response variable Y with change in the predictor variable X.

Order of operations is the rules that determine the sequence in which arithmetic operations should be performed in a numerical expression. The order involves moving from the innermost groups (brackets) to outer groups (brackets), exponents, multiplication and division, and finally, addition and subtraction.

Numerator is the dividend in a fraction. If $\frac{a}{b}$ is a fraction then a is the numerator.

Predictor is a variable that is changed to observe the change in behaviour of another variable. Predictor is also referred to as independent or experimental variable. See also **response**.

Present value (PV) of money is also called the discounted value of money. It is the current value of future cash flow calculated at a discount rate, at a given rate of return.

Primary data is collected first hand. It is original and can be obtained by making a direct contact with the source through interviews, questionnaires, focus groups or observations.

Proportion is an algebraic equation that is used to compare two ratios.

Protractor is a geometrical instrument shaped like a half circle and used for measuring or drawing angles.

Quadratic equation is the equation consisting of one or more variables with the highest exponent 2. It is generally expressed as $ax^2 + bx + c = 0$.

Ratio is a mathematical expression. It is denoted by the symbol : and is used to compare two or more quantities. For example, if the number of boys and girls in a classroom is expressed as 1 : 2, this simply means that for every boy, there are two girls in the class.

Raw data is data that is collected at the source. It gives useful information when it is processed either manually or with computer.

Reciprocal is produced when a number is divided by 1. For example, reciprocal of $x = \frac{1}{x}$

Recurring decimal is a decimal number in which the digits after the decimal point repeat indefinitely.

Relative frequency is the fraction of instances/occurrences observed for a particular value for a variable. For example, if there are 4 occurrences of age 15 years in a sample of 20 people, then relative frequency is $\frac{4}{20}$ or 0.20 or 20%.

Response is a variable that is under observation. Its behaviour is studied in relation to an independent variable. Response is also referred to as dependent or outcome variable. Also see **predictor**.

Sampling error is the magnitude by which the result achieved for the sample deviates from the true characteristics of the population under study.

Secondary data is not original. It has been collected by some earlier research and is now available from published sources such as journals and newspapers.

Simple interest is interest calculated only on the original amount or principal. It is based on specific rate percentage of the principal and calculated for a particular period.

Simultaneous linear equations consist of a pair of equations with two unknown values, x and y. The solution to the equations is possible when the unknown values satisfy both equations. For example, equations $6x + 2y = 56$ and $5x - y = 76$ are simultaneous equations.

Skewness is the asymmetry in a data set. It is observed when data is not distributed uniformly and values tend to be more frequent on one side or other of the mean.

Standard form is a system that expresses very large or small numbers in a more workable form. For example, if sales are $300,000,000, it is written in standard form as $3 × 10^9.

Variable is a measurable or countable data item in statistical analysis. Examples include age, gender, income, quality of service etc. The value of a variable may vary from one unit to another. It may also change over time.

***y*-Intercept** is the value on a graph at which a straight line intersects the *y*-axis.