ISBN 9789976780734

A1 QUANTITATIVE TECHNIQUES

QUANTITATIVE TECHNIQUES

STUDY TEXT

A1

Foundation level

THE NATIONAL BOARD OF ACCOUNTANTS AND AUDITORS TANZANIA (NBAA)

A1
QUANTITATIVE TECHNIQUES

STUDY TEXT

NBAA

NATIONAL BOARD OF ACCOUNTANTS AND AUDITORS

MSEMA KWELI

Edition 1, Version 1

ISBN No 978-9976-78-071-0

Published by

National Board of Accountants and Auditors Mhasibu House, Bibi Titi Mohamed Street, P.O. Box 5128, DAR ES SALAAM Content written by

Get Through Guides Ltd.
Lower First Floor, Bishops House
Market Place Chalfont St. Peter SL9 9HE Buckinghamshire UK

Website: www.GetThroughGuides.com

Email: info@GetThroughGuides.com

The content writer is grateful to The National Board of Accountants and Auditors, Tanzania for permission to reproduce past examination questions. The answers to past examination questions have been prepared by Get Through Guides Ltd.

FOREWORD

The National Board of Accountants and Auditors in Tanzania is a professional body in Tanzania, established under the Auditors and Accountancy Registration Act No 33 of 1972 as amended by Act No 2 of 1995. The Board has been charged among other things, the responsibility to promote, develop and regulate the accounting profession in the country.

In fulfilling its role, NBAA has revised its national accountancy examination scheme and syllabi for students aspiring to sit for Accounting Technician and Professional Examinations. For effective implementation of these syllabi and improve examination results, the Board has prepared study materials for all subjects to assist both examination candidates and trainers in the course of learning and teaching respectively.

The study guides have been prepared in the form of text books with examples and questions to enable the user to have comprehensive understanding of the topics. The study guides cover the wide range of the topics in the syllabi and adequately cover the most comprehensive and complete knowledge base that is required by a leaner to pass the examinations.

These study guides for each subject from ATEC I to final Professional Level will ensure that learners understand all important concepts, know all the workload involved and provide practice they need to do before examinations. The guides have right amount of information with plain language -easy-to-understand, plenty of practice exercises and sample examination questions which are set in a competence based approach.

Competency based study guides have been developed aiming at developing a competent workforce. The guides emphasize on what the individual can do in a workplace after completing a period of training. The training programme therefore is directly related to the expectations of the employer.

These study guides which have been developed under competence based approach are characterized by the following features:-

1. Focus on outcome – The outcomes shown in every topic are relevant to employment industry

2. Greater workplace relevance – the guides emphasize on the importance of applying knowledge to the tasks to be performed at a workplace. This is different from traditional training where the concern has been expressed that theoretical or book knowledge is often emphasized at the expense of the ability to perform the job.

3. Assessments as judgments of competence – The assessment will take into consideration the knowledge, skills and attitudes acquired and the actual performance of the competency.

Study guides are also useful to trainers specifically those who are teaching in the review classes preparing leaners to sit for the professional examinations. They will make use of these study guides together with their additional learning materials from other sources in ensuring that the learners are getting sufficient knowledge and skills not only to enable them pass examinations but make them competent enough to perform effectively in their respectively workplace.

NBAA believes that these standard study guides are about assisting candidates to acquire skills and knowledge so they are able to perform a task to a specified standards. The outcomes to be achieved are clearly stated so that learners know exactly what they have to be able to do, and on the other hand trainers know what training is to be provided and organizations as well know the skills level acquired by their expected accountants.

The unique approach used in the development of these study guides will inspire the learners especially Board's examination candidates to acquire the knowledge and skills they need in their respective examinations and become competent professional accountants in the labor market thereafter.

Pius A. Maneno
Executive Director

STUDY TEXT CONTENTS

A1 – Quantitative Techniques

Features of the book

'The book covers the entire syllabus split into various chapters (referred to as Study Guides in the book). Each chapter discusses the various Learning Outcomes as mentioned in the syllabus.

Contents of each Study Guide

- ‘Get Through Intro’: explains why the particular Study Guide is important through real life examples.
- ‘Learning Outcomes’: on completion of a Study Guide, students will be able to understand all the learning outcomes which are listed under this icon in the Study Guide.

The Learning Outcomes include:

- ‘Definition’: explains the meaning of important terminologies discussed in the learning Outcome.
- ‘Example’: makes easy complex concepts.
- ‘Tip’: helps to understand how to deal with complicated portions.
- ‘Important’: highlights important concepts, formats, Acts, sections, standards, etc.
- ‘Summary’: highlights the key points of the Learning Outcomes.
- ‘Diagram’: facilitates memory retention.
- ‘Test Yourself’: contains questions on the Learning Outcome. It enables students to check whether they have assimilated a particular Learning Outcome.
- Self Examination Questions’: exam standard questions relating to the learning outcomes given at the end of each Study Guide.

EXAMINATION STRUCTURE

The syllabus is assessed by a three hour paper based examination.

The examination will consist of two sections.

Section A One conventional question of 20 marks
Section B Forty multiple choice questions of 2 marks each

CALCULUS

1

Get Through Intro

Calculus is a branch of mathematics in which the derivative is a measure of how a function changes as its input changes. A derivative is how much one quantity varies in response to variations in some other quantity. The process of finding a derivative is called differentiation and the reverse process is called anti-differentiation or integration.

In the context of marginal analysis, in the fields of economics and commerce, differential calculus helps in solving problems related to finding the maximum profit or minimum cost etc. Integral calculus is used to find the cost function when the marginal cost is given; and to find the total revenue when the marginal revenue is given.

Learning Outcomes

a) List the uses of calculus in business.
b) Differentiate a standard form, product of two functions, quotient and function of a function.
c) Determine and identify starting points.
d) Integrate a standard form function.
e) Apply calculus in theory of the firm.
f) Apply the concept of calculus and theory of the firm in accounting and business situations.

1. List the uses of calculus in business.

[Learning Outcome a]

Calculus has two primary branches; differential calculus and integral calculus.

a) Differential calculus studies the variation of a function respective to changes in variables. Derivative measures can be applied to deal with cost minimisation, revenue or profit maximisation, stock market curves, and other functions valuable to the success of any business.

b) Integral calculus, or integration, compacts with areas and volumes of complex figures, e.g. determining the greatest amount of space in a stadium design in order to integrate as many seats as possible.

Business calculus can be applied to determine the equation and graphical shape of a total cost curve, a function's derivative, etc. Derivative is also a financial instrument which plays an important role in the stock market.

1.1 Uses of calculus in various businesses

1. Statistics

In statistics, calculus can be used in assessing survey data to help develop business plans for different companies. Since a survey comprises many different questions with a range of possible answers, calculus allows a more accurate prediction for applicable action.

2. Biologists

In their field, differential calculus is used to derive the exact growth rate in a bacterial culture in a situation where different variables such as temperature and food source are varying. This investigation can help in:

a) either increasing the growth rate of necessary bacteria,

b) or lessening the growth rate of harmful and hypothetically alarming bacteria

3. Credit card companies

In this business, calculus is used to set the minimum payments due on credit card statements. Considering multiple variables such as fluctuating interest rates and a changing available balance, the statement is processed at the exact time.

4. Electrical engineering

Calculus has an advantage of allowing a precise figure to be determined. Here, integration is applied to determine the exact length of power cable needed to connect two substations that are miles apart. This is because the cable is hung from poles and is constantly curving.

5. Architecture

Integration can be used by architects to determine the required amount of materials and labour to construct a curved roof over a new sports stadium. It can also aid in calculating the weight of that curved roof and determine the type of support structure required.

6. Aeronautical engineering

In this field, calculus is frequently used while planning lengthy missions. To launch an experimental probe, they must consider the different revolving velocities of the Earth and the planet the probe is targeted towards, as well as other gravitational influences like the sun and the moon.

Calculus allows each of those variables to be accurately taken into account.

1.2 Conclusion

The above given list is not exhaustive. Obviously, a wide variety of careers regularly use calculus.

This is a more complex branch of mathematics, but touches day to day life in ways too numerous to enumerate.

In today's era, universities, the military, government agencies, airlines, entertainment studios, software companies, and construction companies are some of the employers who seek individuals with a solid knowledge of calculus.

2. Determine and identify starting points.
   Differentiate a standard form, product of two functions, quotient and function of a function.
   [Learning Outcomes b and c]

Definition

Let f be a function that is differentiable at some points belonging in its domain. Then the derivative of f is denoted by f'

Let f(x) be a function, the derivative of f(x) would be:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

The process of finding derivatives is called differentiation.

Example

Let $f(x) = x^4$;

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

$$f'(x) = \lim_{h \to 0} \frac{(x+h)^4 - (x)^4}{h}$$

$$f'(x) = \lim_{h \to 0} \frac{(x^4 + 4x^3h + 6x^2h^2 + 4xh^3 + h^4) - (x)^4}{h}$$

Coefficients and powers of x are calculated on the basis of Pascal's triangle!

$$f'(x) = \lim_{h \to 0} \frac{h(4x^3 + 6x^2h^1 + 4xh^2 + h^3)}{h}$$

$$f'(x) = \lim_{h \to 0} 4x^3 + 6x^2h + 4xh^2 + h^3$$

$$f'(x) = 4x^3$$

Derivatives of constant functions and powers

Rule 1: If k is a constant then: $\frac{d}{dx}k = 0$

Rule 2: if n is any number then: $\frac{d}{dx}x = n\,x^{n-1}$ (hence, if $f(x) = x^5$; $f'(x) = 5x^4$)

Fundamental principles of calculus

If $y = x^n$; $\frac{d}{dx} y = n\, x^{n-1}$

For example, if $y = x^7$, then $\frac{dY}{dx} = 7x^6$

Similarly, if $y = x^6 + 10$, then $\frac{dY}{dx} = 6x^5 + 0 = 6x^5$

Example

If $y = 3\sqrt{x}$ i.e. $3x^{1/2}$

$$\frac{dY}{dx} = 3\left(\frac{1}{2}\right)x^{\frac{1}{2}-1}$$

$$\frac{dY}{dx} = \frac{3}{2}x^{\frac{-1}{2}}$$

$$\frac{dY}{dx} = \frac{3}{2x^{\frac{1}{2}}}$$

$$\frac{dY}{dx} = \frac{3}{2\sqrt{x}}$$

Addition subtraction rule:

If $y = u \pm v$, then:

$$\frac{dY}{dx} = \frac{du}{dx} \pm \frac{dv}{dx}$$

Example

If $y = 12x^3 + 9x^2 - 6x + 4$

$$\frac{dY}{dx} = 36x^2 + 18x - 6$$

Multiplication rule:

If $y = u.v$, then:

$$\frac{dY}{dx} = v.\frac{du}{dx} + u.\frac{dv}{dx}$$

Example

If y = (2x + 3) (5x + 6),

$$\frac{dY}{dx} = v.\frac{du}{dx} + u.\frac{dv}{dx}$$

$$\frac{dY}{dx} = (5x + 6)(2) + (2x + 3)(5)$$
$$= 10x + 12 + 10x + 15$$
$$= 20x + 27$$

Division rule:

If y = u/v, then:

$$\frac{dY}{dx} = \frac{v.\frac{du}{dx} - u.\frac{dv}{dx}}{v^2}$$

Example

If $y = \frac{2x^2 + 13x + 6}{9x + 4}$,

Let, $u = 2x^2 + 13x + 6$ and $v = 9x + 4$

Hence, $\frac{du}{dx} = 4x + 13$ and $\frac{dv}{dx} = 9$

$$\frac{dY}{dx} = \frac{v.\frac{du}{dx} - u.\frac{dv}{dx}}{v^2}$$

$$\frac{dY}{dx} = \frac{(9x+4)(4x + 13) - (2x^2 + 13x + 6)9}{(9x+4)^2}$$

$$\frac{dY}{dx} = \frac{36x^2 + 117x + 16x + 52 - 18x^2 - 117x - 54}{(9x+4)^2}$$

$$\frac{dY}{dx} = \frac{18x^2 + 16x + 42}{(9x+4)^2}$$

$$\frac{dY}{dx} = \frac{2(9x^2 + 8x + 21)}{(9x+4)^2}$$

Chain rule:

If y is a function of u and u is a function of x,

$$\frac{dY}{dx} = \frac{dY}{du} \times \frac{du}{dx}$$

This is a very important rule that leads to the method of substitution which greatly helps us to easily calculate differentiation of composite functions.

Example

If $y = (2x^3 + 9)^4$

Suppose, $u = 2x^3 + 9$; $y = u^4$;

And hence, $\frac{du}{dx} = 6x^2$ and $\frac{dY}{du} = 4u^3$ i.e. $4(2x^3 + 9)^3$

By applying chain rule;

$$\frac{dY}{dx} = \frac{dY}{du} \times \frac{du}{dx}$$

$$\frac{dY}{dx} = 4(2x^3 + 9)^3 (6x^2)$$

$$\frac{dY}{dx} = 24x^2 (2x^3 + 9)^3$$

Derivations of some functions, based on the above principles, are given in the below examples:

Example

1. $y = x^3 + 4x^2 - 6x + 1$

   $$\frac{dY}{dx} = 3x^2 + 8x - 6$$

2. $y = \frac{1}{x} - 10x^3$

   $y = x^{-1} - 10x^3$

   $$\frac{dY}{dx} = -x^2 - 30x^2$$

   $$= \frac{-1}{x^2} - 30x^2$$

3. $y = \log x - 3x^2$

   $$\frac{dY}{dx} = \frac{1}{x} - 6x$$

4. $y = \sqrt{x^2 + 2x + 1}$

   By applying chain rule

   $$\frac{dY}{dx} = \frac{1(2x + 2)}{2\sqrt{x^2 + 2x + 1}}$$

   $$\frac{dY}{dx} = \frac{2(x + 1)}{2\sqrt{x^2 + 2x + 1}}$$

   $$\frac{dY}{dx} = \frac{(x + 1)}{\sqrt{\cancel{(x + 1)(x + 1)}}} = 1$$

Test Yourself 1

Differentiate the function: $y = \frac{4x+5}{\sqrt{2x+7}}$

Test Yourself 2

Differentiate the function: $y = (x^3 + 3x + 9)(2x^2 + 7)$

3. Integrate a standard form function.

[Learning Outcome d]

The process of determining the derivative of the function is called differentiation. The reverse process is known as integration. Converting the derivative to a normal function is called anti-derivation or integration.

For example the differentiation of $y = x^3$ is $3x^2$. The integration of $3x^2$ is $x^3 + c$.

Here, c is a constant

Refer to the below table of integration:

Function f (x)	Indefinite integral Jf (x) dx
Constant 'k'	kx + C
x^n	$\frac{x^{n+1}}{n+1} + C$
1/x (or x-1)	$\|X\| + C$

Example

Find the indefinite integral of $7x^3$.

$\int 7x^3 dx$

$= \frac{7}{4}x^4$

Rules of integration

Integral of f (x) $\pm$ g (x) = $\int f(x)dx \pm \int g(x)dx$

Integral of f (x) x g (x) = $\int f(x)g'(x)dx \pm \int f'(x)g(x)dx$

Rules of integration are similar to the rules of differentiation.

Definite and indefinite integrals

The main distinction between using anti-differentiation when finding a definite versus an indefinite integral is the constant of integration. Indefinite integral is when the integration is done without any limits whereas definite integral is when the integration is done within given limits.

Example

Indefinite integral: $\int f(x)dx$

Definite integral: $\int_a^b f(x)dx$

Let us study some examples of integration

Example

$\int (9x^3 + 24x - 19)dx$

Hence, function y is derived as follows:

$= \frac{9x^{3+1}}{4} + \frac{24x^{1+1}}{2} - 19x + c$

$= \frac{9x^4}{4} + 12x^2 - 19x + c$

Example

$\int \frac{1}{(x+9)^3} dx$

Here, function y is derived as follows:

$= \frac{(x+9)^{-3+1}}{1(-3+1)} + c$

$= \frac{(x+9)^{-2}}{-2} + c$

$= \frac{-1}{2(x+9)^2} + c$

Example

$\int (2x+1)(5x+7)dx$

Here, function y is derived as follows:

$= \int (2x+1)(5x+7)dx$

$= \int (10x^2 + 14x + 5x + 7)dx$

$= \int (10x^2 + 19x + 7)dx$

$= \frac{10x^3}{3} + \frac{19x^2}{2} + 7x + c$

Test Yourself 3

Integrate the following function: $4x^2 + \sqrt{x} + 10$

Tip

How to identify when to use differentiation or integration in your examination.

The following is a basic tip on how to start your example of calculus.

Use differential calculus if your question says:	Use integral calculus if your question says:
Derive the marginal cost / revenue	Find the integral
Find the derivative	Find the area under a curve
Find the rate of change	Find the area between 2 curves
Maximise or minimise the value of 'y'	Find $\int dx$
	Find y when you are given differentiation of y i.e. dY/dx

4. Apply calculus in theory of the firm.
   Apply the concept of calculus and theory of the firm in accounting and business situations.
   [Learning Outcomes e and f]

Calculus is an important branch of mathematics. In mathematics, differentiation is a tool which is extensively used in various fields like physics, circuits' analysis, dynamics etc.

Following are some of the uses of derivatives:

1. Derivatives are useful to find out the maxima and minima of any function or any equations.
2. It is used in many engineering and science problems, especially to know the performance of a moving object.
3. It is used to find out the tangent and normal of the given curve.
4. Differentiation is useful for calculating the velocity and acceleration of a moving particle.
5. It is used to draw a rough curve or to know about the shape of any given function by curve sketching.
6. In numerical analysis, differentiation is used in interpolation; in transform calculus we can use differentiation to solve the given equations.

4.1 Application of differential calculus to calculate optimum selling price and output

1. Revenue and cost functions

Tip

The following functions are used to calculate optimum selling price and output using calculus:

Total cost function

C(x) = F + V(x)
Where, F = fixed costs and V(x) = Variable cost per unit x No. of units

Demand function

x = f (p)
Where, x = number of units demanded and p = selling price per unit

Revenue function

R(x) = x.p
Where, x = number of units sold and p = selling price per unit

Profit function

P(x) = R(x) – C(x)

Break-even point

R(x) = C(x) OR P(x) = 0

The following example would help in understanding the functions and application of calculus in deriving the optimal selling price and output.

Example

Sheni Co's daily cost of producing x number of items of a product is linearly established by C (x) = 420,000x + 14,000,000

(i) If each item is sold for Tshs700,000, the minimum number that must be produced and sold daily to ensure no loss will be calculated as follows:

Here, R(x) = 700,000x and C(x) = 420,000x + 14,000,000
P(x) =700,000x - (420,000x + 14,000,000)
P(x) = 280,000x – 14,000,000
For no profit no loss P(x) = 0
0 = 280,000x – 14,000,000
280,000x = 14,000,000
x = 50 units

Hence, to ensure no loss, the company must produce and sell at least 50 items daily.

(ii) If the selling price is increased by 10%, the revised break-even point would be derived as follows:

When selling price is increased by Tshs70,000 per unit, R(x) = Tshs770,000x and P(x) = 350,000x – 14,000,000
BEP: 350,000x = Tshs14,000,000
Hence, x = 40 units

(iii) If each item is sold for Tshs600,000, the minimum number that must be produced and sold daily to ensure a profit of Tshs4,000,000 will be calculated as follows:

Here, R(x) = 600,000x and C(x) = 420,000x + Tshs14,000,000
P(x) =600,000x - (420,000x + Tshs14,000,000)
Tshs4,000,000 = 180,000x – Tshs14,000,000
180,000x = 18,000,000
x = 100 units

---

2. Average and Marginal Functions

When two variables x and y are related as y = f (x), the average function may be simply defined as $\frac{f(x)}{x}$.

The marginal function is the immediate rate of change of y with respect to x. i.e. $\frac{dy}{dx}$ or $\frac{d}{dx}(f(x))$

(a) Average and marginal costs

(i) Let C = C(x) be the total cost of producing and selling x units of a product, then the average cost (AC) is defined as: AC = $\frac{C(x)}{x}$

Hence, the average cost represents per unit cost.

(ii) Marginal cost (MC) is interpreted as the approximate cost of one additional unit of output. Let C = C(x) be the total cost of producing x units of a product, then the marginal cost (MC), is defined to be the rate of change of C (x) with respect to x

MC = $\frac{dc}{dx}$ or $\frac{d}{dx}(C(x))$

Example

Continuing the previous example of Shiny,

C (x) = 420,000x + Tshs14,000,000

Average cost = $\frac{C(x)}{x} = \frac{420{,}000x + \text{Tshs}14{,}000{,}000}{x}$

Marginal cost = $\frac{dc}{dx}$ = 420,000

(b) Average and marginal revenue

(i) Let R = R(x) be the total revenue from selling x units of a product, then the average revenue (AR) is defined as: AR = $\frac{R(x)}{x}$

Hence, the average revenue represents per unit selling price.

(ii) Marginal revenue (MR) is defined as the rate of change of total revenue with respect to the quantity demanded. MR is interpreted as the approximate revenue received from producing and selling one additional unit of the product.

MR = $\frac{dR}{dx}$ or $\frac{d}{dx}(R(x))$

Example

Continuing the previous example of Shine,

R(x) = 700,000x

Average revenue = $\frac{R(x)}{x} = \frac{700{,}000x}{x}$ = 700,000

Marginal revenue = $\frac{dR}{dx}$ = 700,000

3. Relationship between marginal cost and marginal revenue

a) If marginal revenue is greater than marginal cost, as is the case for small quantities of output, then the firm can increase profit by increasing production. Extra production adds more to revenue than to cost, so profit increases.

b) If marginal revenue is less than marginal cost, as is the case for large quantities of output, then the firm can increase profit by decreasing production. Reducing production reduces revenue less than it reduces cost, so profit increases.

c) If marginal revenue is equal to marginal cost, then the firm cannot increase profit by producing either more or less output. Profit is maximized.

For each unit sold, marginal profit equals marginal revenue minus marginal cost. Then,

a) if marginal revenue is greater than marginal cost, marginal profit is positive
b) if marginal revenue is less than marginal cost, marginal profit is negative
c) if marginal revenue equals marginal cost, marginal profit is zero

Since total profit increases when marginal profit is positive and total profit decreases when marginal profit is negative, it must reach a maximum where marginal profit is zero - or where marginal cost equals marginal revenue. This is because the producer has collected positive profit up until the intersection of MR and MC (where zero profit is collected and any further production will result in negative marginal profit, because MC will be larger than MR).

4. Minimisation of average cost or total cost and maximisation of total revenue, the total profit

(a) Fundamentally, profit maximisation is achieved when dc/dx = dR/dx (or when MC = MR).

(b) Average cost is minimum where, $\frac{d}{dx}(AC) = 0$ and $\frac{d^2}{dx^2}(AC) > 0$

(i.e. first derivation equals to zero and second derivation is greater than zero)

(c) Average revenue is maximum where, $\frac{d}{dx}(R) = 0$ and $\frac{d^2}{dx^2}(R) < 0$

(i.e. first derivation equals to zero and second derivation is smaller than zero)

Example

Rainbow Ltd operates in an entirely different industry. However, it produces to order and carries no inventory. Its demand function can be established as P = 100,000 – 2,000x (where P is the selling price per unit and x is the demanded units in thousands). Total cost function is $C(x) = 1{,}000x^2 + 10{,}000x + 500{,}000$ (where C indicates costs in Tshs'000).

Profit maximisation and sales maximisation are computed as shown below:

(i) Calculation of the total output that will maximise total profits and corresponding profit, selling price per unit and total sales revenue

Revenue function: R(x) = xP

R(x) = x(100,000 – 2,000x)
$R(x) = 100{,}000x - 2{,}000x^2$

Marginal revenue

$MR = \frac{dR}{dx} = 100{,}000 - 4{,}000x$

Total cost function: $C(x) = 1{,}000x^2 + 10{,}000x + 500{,}000$

Marginal cost

$MC = \frac{dc}{dx} = 2{,}000x + 10{,}000$

Optimum output level

MR = MC

100,000 – 4,000x = 2,000x + 10,000
90,000 = 6,000x

x = 15 units (in thousands) i.e. 15,000 units

Selling price per unit at optimum output of 15,000 units

P = 100,000 – 2,000x = 100,000 – (2,000 x 15) = Tshs70,000 per unit

Profit at optimum output of 15,000 units

Profit = TR – TC

= R(x) – C(x)
$= [100{,}000x - 2{,}000x^{2]} - [1{,}000x^2 + 10{,}000x + 500{,}000]$
$= -3{,}000x^2 + 90{,}000x - 500{,}000$
$= -3{,}000(15)^2 + 90{,}000(15) - 500{,}000$
= -675,000 + 1,350,000 – 500,000
= Tshs175,000 (in thousands) i.e. Tshs175,000,000

Continued on the next page

(ii) Calculation of the total output that will maximise total revenue and corresponding profit (or loss), selling price per profit and total sales revenue

Tip

Profit maximisation for sales revenue maximisation; total revenue will be maximised when MR i.e. $\frac{d}{dx}(R) = 0$ and $\frac{d^2}{dx^2}(R) < 0$

MR = $\frac{dR}{dx}$ = 100,000 – 4,000x = 0 (first derivative)

Hence, x = 25 (i.e. 25,000 units)

Selling price per unit at optimum output of 25,000 units

P = 100,000 – 2,000x = 100,000 – (2,000 x 25) = Tshs50,000 per unit

Profit (or loss) at optimum output of 25,000 units
Profit = TR – TC
= R(x) – C(x)
= $[100{,}000x - 2{,}000x^{2]} - [1{,}000x^2 + 10{,}000x + 500{,}000]$
= $-3{,}000x^2 + 90{,}000x - 500{,}000$
= $-3{,}000(25)^2 + 90{,}000(25) - 500{,}000$
= -1,875,000 + 2,250,000 – 500,000
= -125,000 (in thousands) i.e. Loss of Tshs125,000,000

Test Yourself 4

Silver Co manufactures a single product whose variable cost per unit and selling price per unit is Tshs150,000 and Tshs200,000 respectively. It has incurred Tshs100 million as fixed costs.

(a) Establish the functions and compute break-even point.
(b) Determine the profit at the sales volume of 2,500 units.

4.2 Application of integral calculus to calculate optimum selling price and output

In the earlier sections, we studied how marginal cost and revenue functions are derived using differential calculus from total cost and revenue functions. In this section, we will study how total cost and revenue functions are derived from marginal cost and revenue functions using integral calculus.

Example

Suppose C (x) is a cost function and its marginal cost MC = $\frac{dc}{dx}$ = 2,000x + 10,000, using integral calculus, total cost function can be derived as follows:

$\frac{dc}{dx}$ = 2,000x + 10,000

C (x) = $\int 2{,}000X + 10{,}000 \, dx$

C (x) = $\frac{2{,}000X^{1+1}}{2} + 10{,}000X^1 + k$ Here, k is constant.

C (x) = $1{,}000x^2 + 10{,}000x + k$

In the previous example, we learnt how total cost function can be derived. Similarly, from marginal revenue function, total revenue function can be derived.

In case an organisation knows the total cost amount and marginal cost function, using integral calculus, the amount of fixed cost can be derived.

Example

Continuing the previous example of marginal cost,

If the total cost of producing 2,500 units is Tshs9,000 million, the fixed cost can be calculated as follows:
C (x) = $1{,}000x^2 + 10{,}000x + k$

Where,

x = 2,500
C (x) = Tshs9,000 million
Hence,
$9{,}000{,}000{,}000 = 1{,}000\,(2{,}500)^2 + 10{,}000\,(2{,}500) + k$
$9{,}000{,}000{,}000 = 6{,}250{,}000{,}000 + 25{,}000{,}000 + k$
Hence, k = 2,725,0000 i.e. Tshs2,725 million

Test Yourself 5

The marginal revenue function and the marginal cost function of Simba Co are as follows: MR function: $10x^2 – 10x – 20$
MC function: $3x^2 - 300x + 600$

Calculate

(a) The number of units that will generate maximum or minimum revenue.
(b) The total revenue, if 300 units are produced and sold

Answers to Test Yourself

Answer to TY 1

Here, u = 4x + 5 and v = $\sqrt{2x+7}$

$$\frac{du}{dx} = 4 \text{ and } \frac{dv}{dx} = \frac{1}{2\sqrt{2x+7}} \times 2 = \frac{1}{\sqrt{2x+7}}$$

Now by applying division rule:

$$\frac{dY}{dx} = \frac{v.\frac{du}{dx} - u.\frac{dv}{dx}}{v^2}$$

$$\frac{dY}{dx} = \frac{(\sqrt{2x+7}).(4) - (4x+5).\frac{1}{\sqrt{2x+7}}}{\left(\sqrt{2x+7}\right)^2}$$

$$\frac{dY}{dx} = \frac{(2x+7)(4) - 1(4x+5)}{(2x+7)\left(\sqrt{2x+7}\right)}$$

$$\frac{dY}{dx} = \frac{8x + 28 - 4x - 5}{(2x+7)\left(\sqrt{2x+7}\right)}$$

$$\frac{dY}{dx} = \frac{4x + 23}{(2x+7)\left(\sqrt{2x+7}\right)}$$

Answer to TY 2

Here, $u = (x^3 + 3x + 9)$ and $v = (2x^2 + 7)$

Hence, $\frac{du}{dx} = 3x^2 + 3$ and $\frac{dv}{dx} = 4x$

$$\frac{dY}{dx} = v.\frac{du}{dx} + u.\frac{dv}{dx}$$

$$\frac{dY}{dx} = (2x^2 + 7)(3x^2 + 3) + (x^3 + 3x + 9)(4x)$$

$$\frac{dY}{dx} = 6x^4 + 6x^2 + 21x^2 + 21 + 4x^4 + 12x^2 + 36x$$

$$\frac{dY}{dx} = 10x^4 + 39x^2 + 36x + 21$$

Answer to TY 3

$$= \int 4x^2 + x^{\frac{1}{2}} + 10dx$$

$$= \frac{4}{3}x^3 + \frac{2}{3}x^{\frac{3}{2}} + 10x + c$$

$$= \frac{4x^3}{3} + \frac{2}{3\sqrt{x^3}} + 10x + c$$

Answer to TY 4

Let (x) be the number of units sold.

Hence,

Cost function: C(x) = F + V(x) = Tshs100,000,000 + Tshs150,000x
Revenue function: R (x) = x.p = Tshs200,000x
Profit function: P(x) = R(x) – C(x) = Tshs200,000x – (Tshs100,000,000 + Tshs150,000x)

Break-even point: R(x) = C(x)

Hence,

Tshs200,000x	= Tshs100,000,000 + Tshs150,000x
Tshs50,000x	= Tshs100,000,000
x	= 2,000 units

Sales volume beyond 2,000 units will generate profits

Profit function: P(x)
= R(x) – C(x)
= Tshs200,000x – (Tshs100,000,000 + Tshs150,000x)
= (Tshs200,000 x 2,500 units) – [Tshs100,000,000 + (Tshs150,000 x 2,500 units)
= Tshs500,000,000 - Tshs475,000,000
= Tshs25,000,000

Answer to TY 5

(a) Given, $\frac{dR}{dx}$ as $10x^2 - 10x - 20$ is assuming zero (for maximisation of revenue)

$(10x - 20)(10x + 10) = 0$
$(x - 2)(x + 1) = 0$ (dividing with 10 on both sides)
Hence, x = 2 or -1, where -1 is not possible
Hence, x = 2 units

By putting the value of x in second derivation of marginal revenue

$\frac{d^2R}{dx^2} = 20x - 10 = 20(2) - 10 = 30$ which is greater than zero. It signifies that revenue will be minimum at the sales / production of 2 units.

(b) Total revenue when 300 units are produced and sold

$$\int_0^{300} \frac{dR}{dx} dx$$

$$\int_0^{300} 10X^2 - 10X - 20 dx$$

$$\frac{10X^3}{3} - \frac{10X^2}{2} - 20X$$

Here, the lower limit must be 0 and the upper limit must be 300.

$= \frac{10(300)^3}{3} - \frac{10(300)^2}{2} - 20(300)$
$= 90{,}000{,}000 - 450{,}000 - 6{,}000$
$= 89{,}544{,}000$

Self Examination Questions

Question 1

Differentiate: $y = \frac{5}{(3x+2)^2}$

Question 2

Integrate: $y = (9x + 12)^5$

Question 3

Which of the following statements is correct?

A Indefinite integral is when the integration is done without any limits
B Definite integral is when the integration is done within given limits
C Both A and B
D Neither A nor B

Question 4

AB Co produces branded watches. Based on their study of production costs and demand curve, it has determined the following demand and cost functions:

P = 80,000 – 3,000x (demand units are in thousands)
C = $1{,}000x^2$ + 20,000x + 100,000 (cost figures are in thousands)

Required:

(a) Determine the level that maximises sales revenue.
(b) Determine the level that maximises profit.
(c) Determine the optimal price to maximise profits.

Answers to Self Examination Questions

Answer to SEQ 1

$$y = \frac{5}{(3x+2)^2} = \frac{5}{9x^2+12x+4}$$

$$\frac{dY}{dx} = \frac{v.\frac{du}{dx} - u.\frac{dv}{dx}}{v^2}$$

$$= \frac{(9x^2+12x+4)(0) - 5(18x+12)}{(9x^2+12x+4)}$$

$$= \frac{-30(3x+2)}{(3x+2)^2}$$

$$= \frac{-30}{(3x+2)}$$

Answer to SEQ 2

$$\int (9x+12)^5\,dx$$

$$= \frac{(9x+12)^{5+1}}{9(5+1)} + c$$

$$= \frac{(9x+12)^6}{54} + c$$

Answer to SEQ 3

The correct option is C.

Indefinite integral is when the integration is done without any limits i.e. $\int f(x)\,dx$ is an indefinite integral. Definite integral is when the integration is done within given limits.
For example,

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

Answer to SEQ 4

Optimisation of revenue function

R(x) = x.p
= x (80,000 – 3,000x)
= 80,000x – 3,000x^2

Total revenue will be maximised when MR i.e. $\frac{d}{dx}(R) = 0$ and $\frac{d^2}{dx^2}(R) < 0$

Marginal revenue (MR) = 0
80,000 – 6,000x = 0
x = 13.33 = 13,334 watches (approx.)

Determination of second derivative to decide whether x = 13,334 corresponds to optimum value of R

$\frac{d^2}{dx^2}(R) = -6 < 0$

Since, it is negative, x = 13,334 watches corresponds to maximum value of R.

Optimisation of profit function

P(x) = R(x) – C(x)
P(x) = [80,000x – 3,000x^2] – [1,000x^2 + 20,000x + 100,000]
P(x) = -4,000x^2 + 60,000x – 100,000

Total profit will be maximised when MP i.e. $\frac{d}{dx}(P) = 0$ and $\frac{d^2}{dx^2}(P) < 0$

MP = -8,000x + 60,000
0 = -8,000x + 60,000
x = 60,000/8,000
x = 7.5 (i.e. 7,500 watches)

Determination of second derivative to decide whether x = 7,500 corresponds to optimum value of R

$\frac{d^2}{dx^2}(P) = -8 < 0$

Since, it is negative, x = 7,500 watches corresponds to maximum value of R.

Optimal price for profit maximisation

P = 80,000 – 3,000x
= 80,000 – 3 ,000(7.5)
= Tshs57,500

Total revenue at this point will be: 7,500 units x Tshs57,500 = Tshs431,250,000

Total profit at this point will be:
= -4,000x^2 + 6,0000x – 100,000
= -4,000(56.25) + 60,000(7.5) – 100,000
= -225,000 + 450,000 – 100,000
= 125,000 (i.e. Tshs125,000,000)

2

LINEAR REGRESSION AND CORRELATION ANALYSIS

Get Through Intro

In health sciences the relationship between blood pressure and age, consumption level of some nutrient and weight gain, etc. are being studied; the nature and strength of relationship may be examined by statistical techniques, namely, Correlation and Regression analysis. In this Study Guide, we shall discuss correlation and regression analysis.

The study of correlation enables us to evaluate and interpret the relationship between any given values that vary together. It guides us into the analysis of interrelated variables. Regression analysis further helps to determine the extent of changes.

An organisation might be required to assess the correlation between the number of customers in the area and the sales volume. This might help the organisation to plan production based on the extent of correlation between the two.

Learning Outcomes

a) Fit a linear relationship for any two related variables.
b) Estimate unknown values of the dependent variable for given independent variables.
c) Calculate correlation coefficient by both product moment method and rank method.
d) Calculate coefficient of determination.
e) Interpret slope y intercept correlation coefficient and coefficient of determination.
f) Conduct test for slope, and coefficient of correlation.
g) Apply the concept of linear regression and coefficient analysis in accounting and business situations.

1. Fit a linear relationship for any two related variables.
[Learning Outcome a]

A linear relationship between any two related variables can be studied by various mathematical and statistical methods like scatter graph method, correlation analysis, regression analysis, etc.

Correlation refers to the relationship of two variables or more. For example, the relation between height of father and son, crop and rainfall, wage and price index, share and debentures etc. demonstrates the correlation.

Definition

Correlation is statistical analysis which measures and analyses the degree or extent to which the two variables fluctuate with reference to each other. However, it does not indicate cause and effect relationship.

Example

A rise in the temperatures leads to a greater sale of cotton garments. This implies a relationship between temperature and the sale of cotton clothes; but at what temperature how many cotton garments could be sold cannot be predicted by correlation.

However it should be borne in mind that correlation is not always a result of logical reasoning. It might be due to pure chance.

Example

One study observed that the rise in ice cream sales correlates to a rise in crime rates. There is no direct reasoning for this. The high temperatures of summer caused high ice cream sales and hot weather also made more people stay out late at night, which caused the crime rates to rise, as thieves could drop in easily.

Thus correlation only states a relation based on statistics and not essentially on reason.

1.1 Correlation and correlation coefficient

Definition

Correlation signifies a mutual relationship, that is, it is a measure of the extent of interdependence among variable quantities.

This relation between two variables is measured by a coefficient, the ‘correlation coefficient’.

Definition

Correlation coefficient is the factor that represents the correlation between variables. It is denoted by r.

In other words, the coefficient depicts the extent to which the variables are related, whether the relationship is strong (influential), moderate or weak (non-influential). It underlines how a change in the behaviour of one variable influences a change in the behaviour of another variable.

Correlation is a measure of the extent of interdependence in a two-way relationship between two or more variables. When a change in the behaviour of one variable causes some change in the behaviour of the other variable, the variables are said to be correlated.

There can be no relation between two independent variables if a change in the behaviour of one independent variable results in no change in the behaviour of the other. Since correlation can be expressed mathematically, it is represented by a symbol called ‘r’ known as the correlation coefficient.

In other words, a correlation coefficient measures the extent of correlation between two interdependent variables.

Tip

Whenever the correlation between two variables as denoted by r is either exactly or near to -1, 0, or +1, you can easily interpret the nature of relation between the correlation value and the determining factor.

1.2 Linear relationship

The relation is either perfectly positive, perfectly negative or no correlation. The relation between two variables is strongly influenced by each other.

1. Degrees of correlation

Definition

Degrees refer to the extent to which variables are correlated, that is, the extent to which they influence each other's behaviour.

The varying degrees of correlation can be:

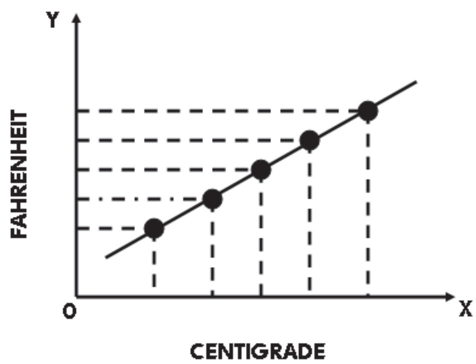
(a) Perfect

In the case of perfect correlation, there is a precise linear relationship between the given variables. The given values for these variables can be plotted on a straight line.

Example

Conversion of temperature expressed as degrees Fahrenheit to degrees centigrade. As the centigrade rises, there is an obvious rise in the Fahrenheit readings.

Plotting this graphically by taking degrees centigrade on X - axis and degrees Fahrenheit on Y - axis, we get



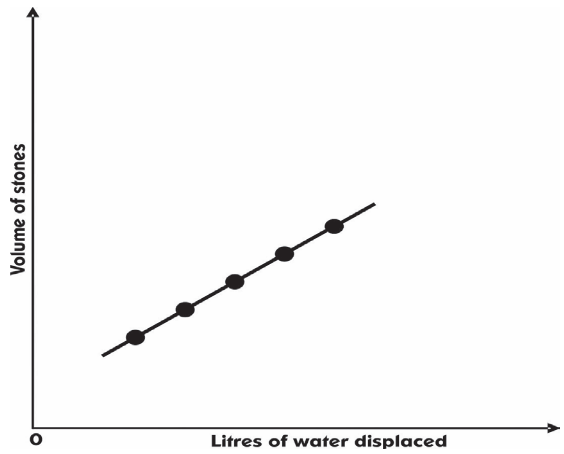
Thus by plotting the points we get a straight line indicating perfect correlation between degrees centigrade and degrees Fahrenheit.

The above graph depicts a relation where both the variables move in the same direction. However, perfectly correlated variables can move in opposite directions. When this is true, as one variable increases the other is bound to decrease.

Example

If you put stones in a tumbler filled with water the level of water will rise in exact proportion to the volume of stones put in.

In this case the ‘volume of stones’ and the ‘litres of water displaced’ will be the variables. Taking the ‘volume of stones’ on Y - axis and ‘litres of water’ on X - axis we can plot the graph as:



As the volume of the stones put in the tumbler increases, the litres of water displaced increases in exact proportion.

(b) Partial

There is no precise linear relationship in partial correlation. These results in some values plotted above the line and others below the line. Here, a change in the behaviour of one variable causes disproportionate change in the value of the other variable.

Example

A car containing 10 litres of petrol will travel further on a smooth road than on one riddled with potholes. As such, mileage of the car increases as the roads become smooth.

In this case we cannot say that the proportionate change in the mileage will always be the same as the change in the quality of the road. Thus, the variables, though correlated, vary disproportionately.

Graphically it will be represented as:

Taking Mileage on Y - axis and Quality of roads on X - axis:

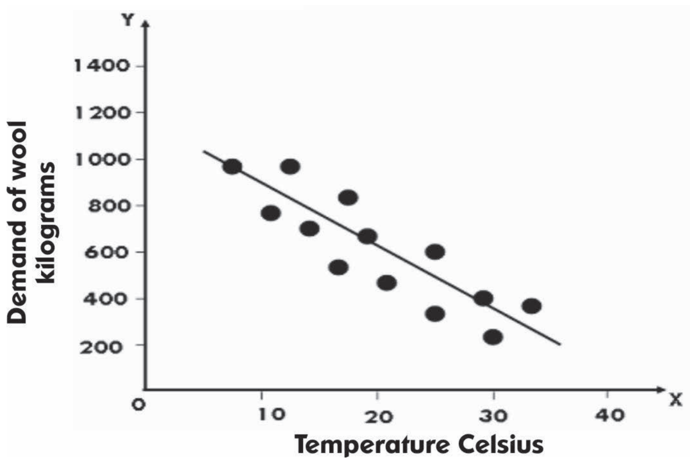
The points plotted are in a definite direction but do not form a straight line. This is because mileage varies disproportionately with the change in the quality of roads. Thus, there is a partial positive correlation.

The relation between two variables can also be partially negative. In that case the variables move in opposite directions.

Example

The demand for wool falls as the temperature increases. It signifies a partial negative correlation as the demand will fall disproportionately with the increase in temperatures.

Taking demand on Y - axis and temperature on X - axis we get the graph as:



As the temperature rises the demand for wool starts decreasing, but disproportionately.

(c) Absent

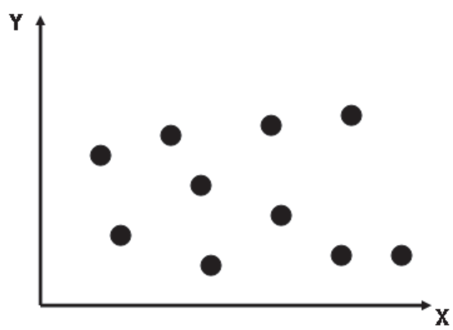
This is the exact opposite of the perfect correlation scenario. There is absolutely no correlation whatsoever between the given variables.

Example

Concrete roads do not affect the amount of rainfall. The number of concrete roads in the area won't affect the amount of rainfall experienced. Thus, this is a case of no correlation or absence of correlation.

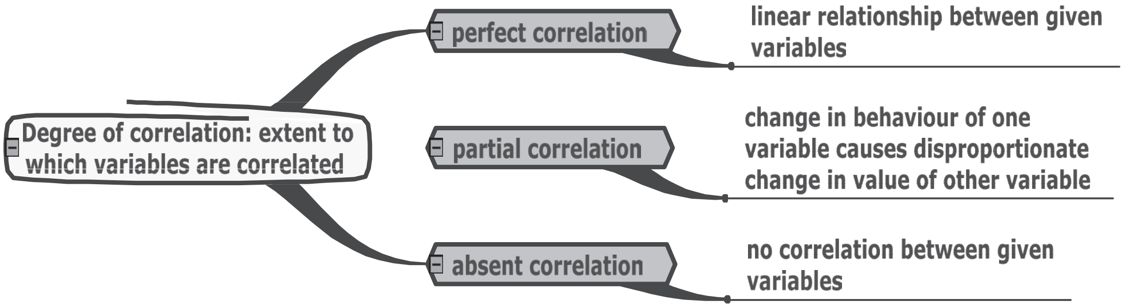
Graphically represented, the diagram for an absent correlation will look as follows:

Absent correlation



Since the variables in an absent correlation are not at all related the points move in random directions making a cluster.

SUMMARY



2. Positive and negative correlation

Correlation is either positive or negative. Irrespective of these aspects, correlation can be partial or perfect.

Important

Correlation can be:

1. Perfectly positive when correlation coefficient is 1
2. Perfectly negative when correlation coefficient is -1
3. Partially positive when correlation coefficient lies between 0 and 1
4. Partially negative when correlation coefficient lies between 0 and -1

(a) Positive correlation

Definition

When the correlated variables move in the same direction the correlation is called positive correlation.

Positive correlation signifies a relationship either between the low or high values of variables. A change in one variable results in a change in another variable in the same direction. In other words, a rise in the value of one variable leads to a rise in the value of the other variable and vice versa. For example a change in a child's age co-varies with his height; the older the child, the taller he is.

Example

As the altitude at which an area is located increases, the snowfall experienced by the area also increases. In this case as the value of one variable – 'altitude' increases, the value of the other variable – 'snowfall' also increases. This is an example of a positive correlation.

(b) Negative correlation

Definition

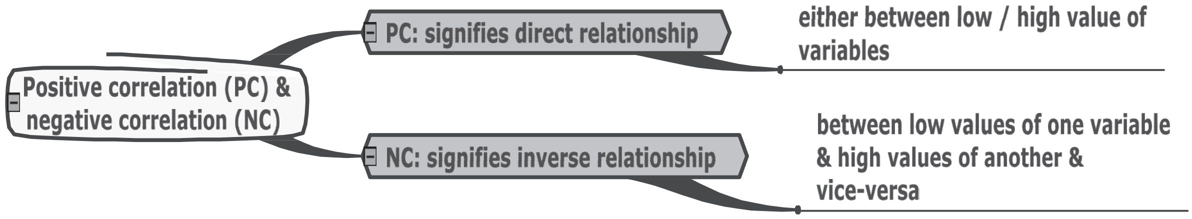
When the correlated variables move in opposite directions the correlation is called negative correlation.

Negative correlation signifies an adverse relationship, one where the low values of one variable and high values of the other occur at the same time. This implies an inverse relationship, a situation where things vary inversely.

Example

Considering the high altitudes once again we can say that higher altitudes are correlated with lower temperatures. Thus in this case the higher magnitudes of one variable lead to lower magnitudes of the other. This is a case of inverse correlation.

SUMMARY



Test Yourself 1

Which of the following is the property of correlation coefficient?

A Correlation coefficient is the factor that represents the correlation between two variables
B Correlation coefficient depicts the extent to which the variables are related
C Correlation coefficient underlines the influence of change of behaviour in one variable on another
D Correlation coefficient exhibits all the above properties

Test Yourself 2

Degrees of correlation between variables can be:

A Perfect
B Partial
C Absent
D All perfect, partial and absent

Test Yourself 3

Which of the following is correct in regards to a positive correlation?

A Higher magnitudes of one variable correspond with higher magnitudes of the other
B Higher magnitudes of one variable correspond with lower magnitudes of the other
C Lower magnitudes of one variable correspond with higher magnitudes of the other
D None of the above

2. Calculate correlation coefficient by both product moment method and rank method. [Learning Outcome c]

2.1 Correlation coefficient by product moment method

The most common method of measuring the degree of correlation is the Karl Pearson correlation coefficient method, which is sensitive only to a linear relationship between two variables (which may exist even if one is a nonlinear function of the other). This method is also known as product moment method.

The various formulae for calculating ‘r’ under Karl Person’s method are:

$$r_{xy} = \frac{\text{Covariance}}{\text{Standard deviation}}$$

$$r = \frac{\text{Cov}(x, y)}{ax.ay} \text{ OR } r = \frac{S_{xy}}{S_x.S_y}$$

Where,

Covariance of x, y: $\text{Cov}(x, y) = \frac{L(X - \bar{X})(Y - \bar{Y})}{n}$

ơ, ơ x y (a_x, a_y) are S.D of x and y respectively

S_{xy} indicates the strength and direction of a linear relationship between X and Y. A covariance is an indicator of the extent to which X and Y have a linear relationship. It doesn't tell us the exact position of the line in the coordinate system. For that, we must use regression.

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{n.a.ax.}$$ where x bar is the mean of x; and y bar is the mean of y

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

Where: X and Y denote data variables for which data is provided;
n = the number of pairs of data for the given variables, used in the analysis.

Tip

The correlation coefficient is measured on a scale that varies from + 1 through 0 to – 1. Thus, correlation coefficient, r, must have a value between -1 and +1. If the value goes beyond these boundaries, then there is an error in calculation. These lower and upper value limits signify the following:

- Where r = -1: correlation is perfectly negative
- Where r = +1: correlation is perfectly positive
- Where r = 0: correlation is absent
- Where r = any value which is between -1 to +1 (except 0): correlation is partially positive or negative

To have a better understanding let us try to calculate the coefficient of correlation:

Example

The following data is given regarding production of 100ml perfume bottles for a 10 year period and the cost of their manufacture. This cost varies with the number of bottles produced. Compute the correlation coefficient.

Years	Production ('000 Bottles)	Cost (Tshs'000)
1	25	275
2	28	300
3	30	333
4	33	350
5	35	377
6	37	400
7	40	444
8	43	475
9	46	500
10	50	533

In the above information, production (volume of bottles) and the associated 'costs' are the variables that are correlated. Let us denote 'volume of bottles' as X and 'costs' as Y. In the above table, the costs are increasing as the volume of production increases. Change in one variable causes change in the other.
To calculate the value of r – the correlation coefficient, we have to first assess whether we have all the required data for the formula:

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

Continued on the next page

We now proceed to calculate the values for the above components of the equation with the help of the given information.

X (Bottles)	Y (Cost)	XY	X^2	Y^2
25	275	6,875	625	75,625
28	300	8,400	784	90,000
30	333	9,990	900	110,889
33	350	11,550	1,089	122,500
35	377	13,195	1,225	142,129
37	400	14,800	1,369	160,000
40	444	17,760	1,600	197,136
43	475	20,425	1,849	225,625
46	500	23,000	2,116	250,000
50	533	26,650	2,500	284,089
$\Sigma X = 367$	$\Sigma Y = 3{,}987$	$\Sigma XY = 152{,}645$	$\Sigma X^2 = 14{,}057$	$\Sigma Y^2 = 1{,}657{,}993$

$(\Sigma X)^2 = (367)^2 = 134{,}689$; $(\Sigma Y)^2 = (3{,}987)^2 = 15{,}896{,}169$; n = pairs of data = 10.

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

$$\frac{(10 \times 152{,}645) - (367 \times 3{,}987)}{\sqrt{\left((10 \times 14{,}057) - (134{,}689)\right)\left((10 \times 1{,}657{,}993) - (15{,}896{,}169)\right)}}$$

$$\frac{1{,}526{,}450 - 1{,}463{,}229}{\sqrt{(140{,}570 - 134{,}689)(16{,}579{,}930 - 15{,}896{,}169)}}$$

$$\frac{63{,}221}{\sqrt{(5{,}881)(683{,}761)}}$$

$$\frac{63{,}221}{\sqrt{63{,}413}}$$

= 0.997

This value signifies partial positive correlation between the volume of perfume bottles and manufacturing costs. It is partial as the answer is not '1'. The figure '1' denotes a perfect relationship as explained earlier. However, since this figure very closely approximates '1' we may say that it is extremely closely correlated.

Correlation in a time series: a time series is a given period of time for which data is collected and recorded. Correlation can exist in a time series when there is a relationship between the 'period of time' and the 'values' recorded for that time period.

Example

The demand for the product of Tanga Cement Co. has gradually declined over the years due to similar better quality
products in the market. The data for this product is given below. Ascertain whether the production figures show a trend, that is, whether there exists any correlation between the year and production volume.

Years	Production ('000 units)
0	40
1	36
2	30
3	28
4	22

Answer

Here the 'years' and the production units are the variables. Let 'X' denote the 'years' and 'Y' denote the 'production units'.

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

X	Y	XY	X^2	Y^2
0	40	0	0	1,600
1	36	36	1	1,296
2	30	60	4	900
3	28	84	9	784
4	22	88	16	484
$\Sigma X = 10$	$\Sigma Y = 156$	$\Sigma XY = 268$	$\Sigma X^2 = 30$	$\Sigma Y^2 = 5{,}064$

$$= \frac{(5 \times 268) - (10 \times 156)}{\sqrt{((5 \times 30) - 100)((5 \times 5{,}064) - 24{,}336)}}$$

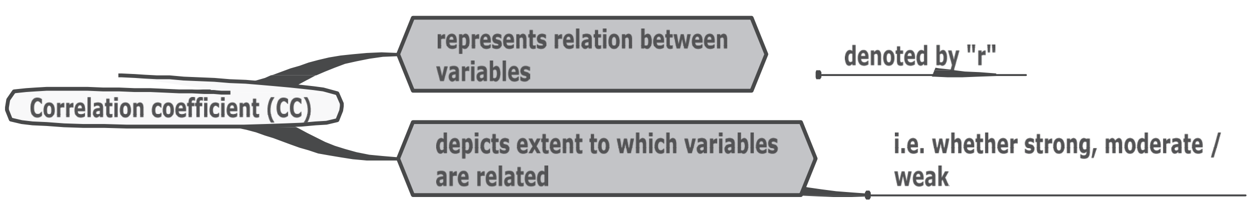
$$= \frac{1{,}340 - 1{,}560}{\sqrt{(150 - 100)(25{,}320 - 24{,}336)}}$$

$$= \frac{-220}{\sqrt{50 \times 984}}$$

$= -0.991$

Thus there is partial negative correlation between the years of production and the units produced. However, since this figure very closely approximates '-1' we may say there is almost perfect negative correlation between them.

SUMMARY



2.2 Properties of correlation

1. The correlation coefficient always lies between –1 and +1
2. 'r' is independent of change of origin and scale. Mathematically, if X and Y are given and they are transformed to the new variables u and v by the change of origin and scale viz, u = (x – A)/h and v = (y – B) /k ; h >0, k >0 Where A, B, h >0, k >0; then the correlation coefficient between x and y is the same as the correlation coefficient between u and v, i.e. r (x,y) = r (u, v) $r_{xy} = r_{uv}$ (Refer to the example of Charles below)
3. It is a pure number, independent of units of measurement.
4. Independent variables are uncorrelated, but the opposite is not true. If r = 0, the variables are uncorrelated. In other words, there is no linear (straight line) relationship between the variables. However, r = 0 does not imply that the variables are totally independent.
5. The correlation coefficient is the geometric mean of two regression coefficients.
6. The correlation coefficient of two variables x and y is symmetric; $r_{xy} = r_{yx}$.

Example

Charles is engaged in the production of leather belts and purses. The following data relates to the total cost incurred and man hours utilised for the production of leather belts and purses.

Months	Jan	Feb	Mar	Apr	May	Jun
Man hours	60	55	45	90	105	45
Total Costs (Tshs million)	748	649	510	900	1000	650

From the above data, correlation can be derived using the following formulae:

$$r = \frac{n\Sigma uv - \Sigma u \Sigma v}{\sqrt{\left(n\Sigma u^2 - (\Sigma u)^2\right)\left(n\Sigma v^2 - (\Sigma v)^2\right)}}$$

Where, u = (X – A)/h and v = (y – B)

Assuming, A = 50; h = 5 and B = 750

> The value of A, h and B can be assumed on the bases of the data provided. For example, here each man hour can be divided by 5; we have chosen h as 5 and since $\bar{X}$ is 66.67, we have assumed A as 50 (nearest number which is easy to subtract from the given man hours).
>
> Similarly, $\bar{Y}$ is 742.8; we have assumed B as 750.

u	v	uv	u^2	v^2
2	-2	-4	4	4
1	-101	-101	1	10201
-1	-240	240	1	57600
8	150	1200	64	22500
11	250	2750	121	62500
-1	-100	100	1	10000
20	-43	4185	192	162805
u	v	Uv	u^2	v^2

Tip

Note: we have assumed A as 50, which is smaller than $\bar{X}$, hence Σu is a positive figure. We have assumed B as 750 which is greater than $\bar{Y}$, hence Σv is a negative figure.

Continued on the next page

Calculation of correlation coefficient

$$r = \frac{n\Sigma uv - \Sigma u \Sigma v}{\sqrt{\left(n\Sigma u^2 - (\Sigma u)^2\right)\left(n\Sigma v^2 - (\Sigma v)^2\right)}}$$

$$= \frac{(6 \times 4185) - (20 \times -43)}{\sqrt{\left((6 \times 192) - (400)\right)\left((6 \times 162805) - (1849)\right)}}$$

$$= \frac{25110 + 860}{\sqrt{752 \times 974981}}$$

$$= \frac{25970}{\sqrt{733185712}}$$

= 25970/27077.40

= 0.96 (strong positive correlation)

2.3 Interpretation of correlation coefficients

Interpretation of the coefficients of correlation means trying to understand what they signify. The coefficient 'r' can be used to understand the kind of relation one variable has with the other.

The different types of degrees of correlation imply various interpretations.

(a) Partial positive correlation

Example

Referring to the previous example relating to the production of perfume bottles, a positive correlation coefficient of 0.997 (almost 1) signifies a partial positive correlation between the quantity of perfume bottles manufactured and their manufacturing costs.

The number of bottles and their cost vary in direct proportion. This would imply that a low quantity would correspond with low cost and a higher quantity would correspond with higher costs. This interpretation can help us plan production accordingly making use of the correlation coefficient.

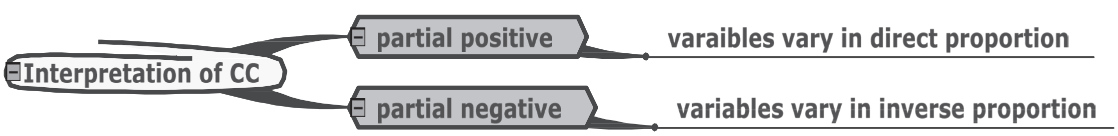
(b) Partial negative correlation

Example

In the example of "Tanga Cement Co" the correlation r was - 0.991. This signifies a partial negative correlation between the years and the units produced.

Expressed differently, years and quantity vary in an inverse proportion. This would imply that every succeeding year would correspond to successive lower production levels. Conversely, successive higher production levels would correspond to preceding years. These statistics will help us take decisions on whether to continue production.

SUMMARY



Standard error (SE)

Standard error of the coefficient of correlation is used for ascertaining the probable error of the coefficient of correlation.

$$SE = \frac{1 - r^2}{\sqrt{N}}$$

Where,
r = Coefficient of correlation
N = No. of pairs of observations

Probable error (PE)

The probable error of coefficient of correlation is an amount which, if added to and subtracted from the value of r, gives the upper and lower limits with in which coefficients of correlation in the population can be expected to lie.

It is 0.6745 times of standard error. PE is used for determining the reliability of the value of r in so far as it depends on the condition of random sampling.

1. If |r |< 6 PE; the value of r is not at all significant. There is no evidence of correlation.
2. If |r | >6 PE; the value of r is significant. There is evidence of correlation.

Example

In a school, there are 2,000 students. Out of them, random 100 students were selected and the correlation between their marks in English and Accountancy was derived as -0.25.

Here,

$$SE = \frac{1 - r^2}{\sqrt{N}} = \frac{1 - 0.0625}{\sqrt{100}} = 0.09375$$

PE = SE x 0.6745 = 0.09375 x 0.6745 = 0.06323

Let's verify the evidence of correlation:

6PE = 6 x 0.06323 = 0.3794

Here, 0.25 < 0.3794 i.e. |r |< 6 PE; the value of r is not at all significant. Therefore, there is no evidence of correlation.

2.4 Rank correlation

The product moment correlation coefficient is used to measure the strength of the linear relationship between two variables, i.e. how close the points on a scatter graph lie to a straight line. It is most appropriate when the points on a scatter graph have an indirect pattern. However, this method is less appropriate when the points on a scatter graph seem to follow a curve or when there are outliers (or inconsistent values) on the graph.

The Spearman's rank-order correlation (denoted as r_s) is the nonparametric version of the Pearson product-moment correlation. Spearman's correlation coefficient measures the strength of association between two ranked variables. It is useful to study the qualitative measure of attributes like honesty, colour, beauty, intelligence, character, morality etc.

Ranking follows in the circumstances where either there is a lack of time and money or suitable measurements may be impossible to quantify the items. In dealing with a correlation problem where the values are in ranks, rank correlation methods are used.

Tip

Product moment correlation coefficient finds out how close two sets of points lie to a straight line or line of best fit. Spearman's rank finds out how closely two sets of points agree with each other.

Method of calculating Spearman's rank correlation:

Step 1: Rank both sets of data (from the highest to the lowest)

Note: there may be two or more items having equal values. In such a case, the same rank is to be given. The ranking is said to be tied; in such circumstances an average rank is to be given to each individual item.

Step 2: Calculate the difference between ranks given for variables x and y
Step 3: Derive the squares of each rank's differences
Step 4: Apply the formula of Spearman's rank correlation

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

Example

Consider the marks obtained in English and Maths by 10 candidates.

Candidates	A	B	C	D	E	F	G	H	I	J
English	66	68	56	75	80	60	55	60	48	74
Maths	84	80	79	55	50	60	70	58	64	77

Spearman's rank correlation can be derived as follows:

X	Y	Rx	Ry	d	d x d
66	84	5	1	4	16
68	80	4	2	2	4
56	79	8	3	5	25
75	55	2	9	-7	49
80	50	1	10	-9	81
60	60	6.5	7	-0.5	0.25
55	70	9	5	4	16
60	58	6.5	8	-1.5	2.25
48	64	10	6	4	16
74	77	3	4	-1	1
				0	210.5

Both candidates F and H have scored 60 marks, hence rank would be: $\frac{\text{6th rank + 7th rank}}{\text{2 candidates}} = 6.5^{th}$ rank

Σd is always equal to zero

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

$$= 1 - \frac{6\,(210.5)}{10(100 - 1)}$$

$$= 1 - \frac{1263}{990}$$

$= 1 - 1.276$

$= -0.276$ (partial negative correlation)

It indicates that students who score good marks in English tend to score comparatively poor marks in Maths (and vice-versa).

The same example is shown under the product moment method as follows:

Example

X	Y	X x X	Y x Y	X x Y
66	84	4356	7056	5544
68	80	4624	6400	5440
56	79	3136	6241	4424
75	55	5625	3025	4125
80	50	6400	2500	4000
60	60	3600	3600	3600
55	70	3025	4900	3850
60	58	3600	3364	3480
48	64	2304	4096	3072
74	77	5476	5929	5698
642	677	42146	47111	43233

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

$$r = \frac{10\,(43233) - (642)\,(677)}{\sqrt{\left(10\,(42146) - (642)^2\right)\left(10\,(47111) - (677)^2\right)}}$$

$$r = \frac{432330 - 434634}{\sqrt{(421460 - 412164)\,(471110 - 458329)}}$$

$$r = \frac{-2304}{\sqrt{(9296)\,(12781)}}$$

$$r = \frac{-2304}{10900.099}$$

r = -0.21

When to use rank method and when to use Pearson's method to derive correlation

Important

(a) If the given data has already been ranked, you should compulsorily use the rank correlation coefficient (R) method.

(b) Where actual values (i.e. not just ranks) of x and y are given, usually, Pearson's coefficient (r) should be used since information is lost when values are converted into their ranks.

(c) Particularly, Karl Pearson's coefficient must be used if one intends to use regression for forecasting (see later in Learning Outcome 3).

Test Yourself 4

From the following data collected in a survey conducted by an advertising firm, find the correlation between the age of the drivers and the number of accidents relating to the drivers in a period of 6 months.

Age	25	30	45	50	55	60	65
No. of accidents	2	2	4	5	6	7	8

3. Apply the concept of linear regression and coefficient analysis in accounting and business situations.
   Estimate unknown values of the dependent variable for given independent variables.
   [Learning Outcomes b and g]

3.1 Correlation and regression analysis

Correlation estimates the degree of association between the variables and makes no a priori assumption as to whether one variable is dependent on the other(s) and is not concerned with the relationship between variables. In other words, it can be said that correlation analysis checks for interdependence of the variables.

On the other hand, regression attempts to describe the dependence of a variable on one (or more) explanatory variables; it indirectly assumes that there is a one-way causal effect from the explanatory variable(s) to the response variable, regardless of whether the path of effect is direct or indirect.

Using the regression equation, the value of the dependent variable may be estimated from the given independent variable.

Example

Study two variables: BMI (Kg/m^2) of pregnant mothers and the birth-weight (BW in Kg) of their new-borns. While correlation analysis would show a high degree of association between these two variables, regression analysis would be able to demonstrate the dependence of a new-born's birth weight on the BMI of the pregnant mother.

However, careless use of regression analysis could also demonstrate that the BMI of a pregnant mother is dependent on the new-born's birth weight - this would indicate that the low birth weight of a new-born causes low birth weight in the mother!

Linear functions using regression analysis

Important terms

1. Linear

   (a) That which is arranged in or extended along a straight line
   (b) Consisting of lines or outlines
   (c) Involving one dimension only
   (d) Progressing from one stage to another in a series of steps

2. Function: it denotes a relationship between one element and another, or between several elements.

3. Regression analysis: a mathematical technique used to:

   (a) Measure relationships between variables based on models
   (b) Establish the magnitude or extent of relationships between these variables
   (c) Make predictions based on these models

3.2 Regression analysis

Regression analysis establishes a relationship between:

a) One or more response variables (also known as dependent variables, explained variables or predicted variables) and

b) Predictors (also called independent variables, explanatory variables or control variables).

After learning about a close correlation between two variables, we will now try to assess the value of one variable given the value of the other. Regression Analysis is a statistical device used for predicting these unknown values of one variable (dependant variable) from the known values of the other (independent variable). This is accomplished through the regression line that describes the average relationship between two variables, say x and y. The regression line is the line which best represents the data on a graph.

For the purpose of this syllabus requirement, you will be dealing with simple linear regression, that is, regression on two variables only.

This is a statistical method, which illustrates the relationship between two variables using a linear equation. It assumes the best estimate of a response.

3.3 Least Squares Method of linear regression analysis

This method envisages plotting the “line of best fit” (the line that best represents the given situation) for a given number of observations, by using a mathematical equation. Here we will make use of the linear equation. The equation portrays a linear relationship between two variables x and y as below:

$$\boxed{y = a + bx}$$

In any given situation we will have the data for only one variable and the other variable will have to be derived using the formula “y = a + bx”. When we have the value of one variable, say x, we now need to have the values of the constants ‘a’ and ‘b’ which can be put in the equation to arrive at the value of y.

The steps mentioned below guide us to derive the equations that will help us find the values of a and b.

The equation “y = a + bx” can be expressed for every period of activity as below:
Period 1: $y_1 = a + bx_1$
Period 2: $y_2 = a + bx_2$
Period n: $y_n = a + bx_n$

If you add these values, you get:
$\Sigma y = na + b\Sigma x$ ………. (1)

Therefore,

$$a = \frac{\Sigma y}{n} - \frac{b\Sigma x}{n}$$

Again, by multiplying both sides of the equation y = a + bx by x, you get:

Period 1: $x_1y_1 = ax_1 + bx_1^2$
Period 2: $x_2y_2 = ax_2 + bx_2^2$
Period n: $x_ny_n = ax_n + bx_n^2$
When these are added together:
$\Sigma xy = a\Sigma x + b\Sigma x^2$ ………… (2)

Substituting the value of ‘a’ as derived in the above equation in equation (2) we get

$$\Sigma xy = \left[\frac{\Sigma y}{n} - \frac{b\Sigma x}{n}\right]\Sigma x + b\Sigma x^2$$

$$\Sigma xy - b\Sigma x^2 = \frac{\Sigma x\Sigma y - b(\Sigma x)^2}{n}$$

$$n(\Sigma xy - b\Sigma x^2) = \Sigma x\Sigma y - b(\Sigma x)^2$$

$n\Sigma xy - nb\Sigma x^2 = \Sigma x\Sigma y - b(\Sigma x)^2$

$n\Sigma xy - \Sigma x\Sigma y = nb\Sigma x^2 - b(\Sigma x)^2$

$n\Sigma xy - \Sigma x\Sigma y = b\,[n\Sigma x^2 – (\Sigma x)^2]$

$$b = \frac{n\Sigma xy - \Sigma x\Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

By using these 2 equations (1 and 2), you can arrive at the value of constants a and b and determine the trend in the cost line accordingly. Here are some noteable points:

a) The straight line equation is $y = a + bx$. In this equation, a and b are constants.

b) Once the values of constants a and b are known, the value of y corresponding to any value of x can be computed.

c) ‘b’ is considered as slope and ‘a’ is considered as Y intercept.

Example

You are provided with the following data:

Month	Production (Units)	Semi-variable overheads (Tshs’000)
July	220	22
August	220	23
September	240	23
October	240	25
November	260	25
December	260	27

Find the regression line by the least squares method. What would be the semi-variable overheads in January if production level increases to 2,700 units?

Answer

Calculation of the information required according to the formula to compute the regression coefficient:

Month	Production volume (x)	Semi-variable overheads (y)	x^2	xy
July	220	22	48,400	4,840
August	220	23	48,400	5.060
September	240	23	57,600	5,520
October	240	25	57,600	6,000
November	260	25	67,600	6,500
December	260	27	67,600	7,020
Total	Σx = 1,440	Σy = 145	Σx^2 = 347,200	Σxy = 34,940

n = 6

You know that:
$\Sigma y = na + b\Sigma x$ ………………………….. (i)
$\Sigma xy = a\Sigma x + b\Sigma x^2$………………………… (ii)

Substituting these values in the above equations:
145 = 6a + 1,440b………………………… (iii)
34,940 = 1,440a + 347,200b…………….. (iv)

Continued on the next page

You will have to solve the above two equations to find the values of a and b. For this, we will have to subtract one equation from the other in such a way, that either ‘a’ or ‘b’ gets cancelled.

Observing equation (iii) we can see that if we multiply 6a by 240 we get 1440a.

So, multiplying equation (iii) by 240:
34,800 = 1,440a + 345,600b……………… (v)

Now deduct equation (v) from equation (iv) to get:

34,940	= 1,440a + 347,200b
(-) 34,800	= 1,440a + 345,600b
140	= 0 + 1,600b
B	= 140/1,600
B	= 0.0875

Substituting this value of b in equation (iii):

145 = 6a + (1,440 x 0.0875)
145 = 6a + 126
6a = 19
a = 3.167

Equation of a straight line is: y = a + bx

By substituting a and b by their values, the equation becomes:
y = 3.167 + 0.0875x

This is the regression line of best fit, where:
y = semi-variable overhead; x = volume of production

If x = 2700,

Then: y = 3.167 + 0.0875x
y = 3.167+ (0.0875 x 2700)
y = 236.25 + 3.167
y = 239.417

Thus, if production level is 2,700 units in January, semi-variable overhead will be Tshs239,
417.

In this example, regression formula is: Y = 3.167 + 0.0875 x; where slope is 0.0875 and Y intercept is 3.167.

a) A slope of 0.0875 means that semi-variable overhead increases by 0.0875 (less than 1) for every 1 unit increase of production.

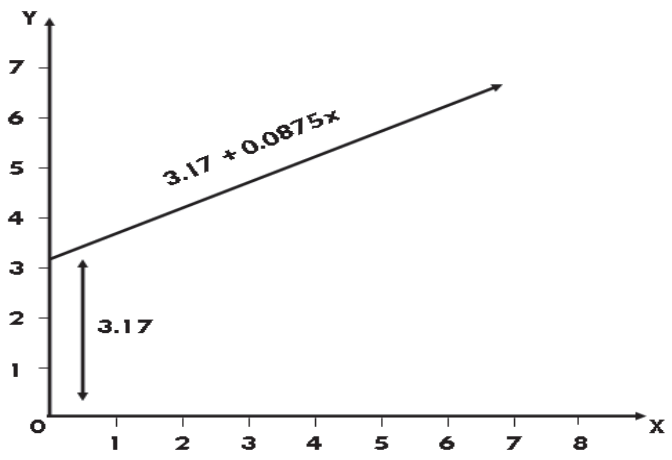
b) The Y intercept means that the regression line crosses the Y axis at Y = 3.167

3.4 Interpretation

The above example portrays a linear relationship between the two variables, units of production and the semi-variable overheads associated with them. Paper A4 tells you that semi-variable overheads are a type of hybrid cost that contains elements of fixed and variable cost. In this case, per unit variable cost remains the same throughout but the fixed costs increase when the output exceeds a specified limit. Therefore these overheads do not rise proportionately to the rise in output.

The above relationship is expressed in the form of a “line of best fit”, which is expressed as a mathematical equation. According to the given example, variable cost per unit will remain constant at Tshs8.75 irrespective of output level. On the other hand, fixed cost “a” will remain constant in total and keep changing at the per unit front depending on the change in output.

The graphical representation of the above equation y = 3.17 + 0.0875x is given below. It is similar to the graphical representation of semi-variable costs that consist of a fixed portion and a variable portion.



3.5 Limitations of the "line of best fit"

1. It is assumed that there is an on-going stable relationship between two variables. The assumption is that x varies only because of variation in y. However, the reason for the variation in x could be different.

Example

The demand for ice creams increases with an increase in temperature. If the demand for ice creams in the current year has increased phenomenally from the previous year, the increase in demand cannot be attributed solely to the high temperatures especially if the temperature level is the same as that of the previous year. The demand can also rise because of a rise in the number of customers.

---

2. Regression analysis assumes relationships existing in the past between the variables will hold good for future periods. The passage of time will lead these relations to vary due to changing environments or altering behaviours of the variables themselves.

Example

To communicate with people living far away, writing physical letters and mailing them through the postal department was the traditional means of communication. However, with the advent of technology, e-mailing has substantially decreased the need to write physical letters. Accordingly, there has been a huge drop in the number of letters physically written per year.

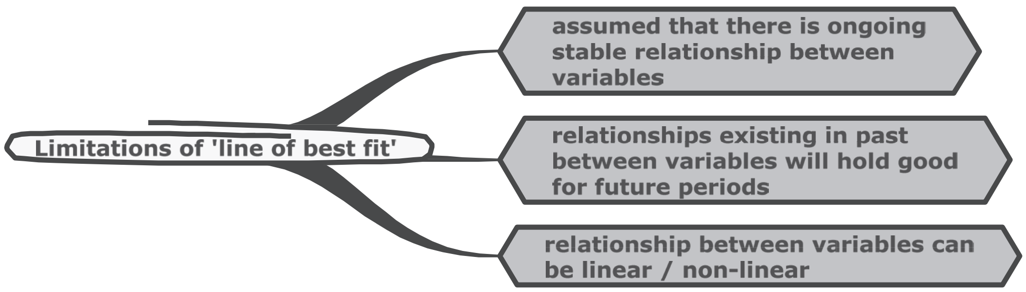
---

3. It presumes a linear relation between variables but the relationship can be nonlinear also. It is possible that the variables do not satisfy the equation "y = a + bx".

Example

In the above two examples (ice cream and temperature, physical letters and mailing) the relationship is not linear as the variables do not vary perfectly with each other. They vary disproportionately with each other. It depicts a curvilinear relationship. Most relationships between variables do not satisfy linearity.

SUMMARY



Test Yourself 5

Which of the following are correct with regard to regression analysis?

(i) In regression analysis the n stands for the number of pairs of data
(ii) �x^2 is not the same calculation as (�x) ^{2}
(iii) �xy is calculated by multiplying the total value of x and the total value of y

A (i) and (ii) only
B (i) and (iii) only
C (ii) and (iii) only
D (i), (ii) and (iii)

Test Yourself 6

In regression line y = a + bx, y stands for yield and x stands for units. a and b are fixed costs and per unit variable costs respectively. a is Tshs200,000 and b is Tshs2,500. If 350 units are produced, the costs will be:

A 1,075,000
B 200,000
C 875,000
D 1,100,000

Test Yourself 7

In regression line y = a + bx, y denominates the telephone bill amount and x denominates the units consumed. When 200 units were consumed, the bill amount was Tshs2,500,000 and when 500 units were consumed, the bill amount was Tshs4,900,000. Hence a and b are respectively:

A Tshs500,000 and Tshs10,000
B Tshs900,000 and Tshs9,000
C Tshs600,000 and Tshs15,000
D None of the above

Interpretation of b

Since regression coefficient (b) is the slope of the regression line, its magnitude gives an indication of the steepness of the line and it can be interpreted as follows:

(i) If b = 0, it means the line is parallel to the x axis.
(ii) If b is high and positive, it gives a very steep and upward sloping regression.
(iii) If b is negative, it gives a downward sloping regression line.

The regression line can be used for prediction. When the regression equation or regression line is used to obtain the y value corresponding to a given x value, we say that x is used to predict y. We can similarly use y to predict x.

3.6 Use of linear regression analysis in the analysis of cost data

1. In order to overcome the shortcomings of the high / low method of analysing the costs into its fixed and variable components, can be done more accurately by the statistical method of regression analysis (also known as least squares).

2. An organisation needs to estimate the future production, resource consumption, sales etc. for budgeting purposes. The estimation is made by identifying the dependent and independent variables For example, demand for a particular product (dependent variable) depends on its price (independent variable) and the price of the product depends on the production costs and other administrative and selling overheads. Regression analysis is a mathematical tool which describes and evaluates the relationship between two or more variables.

3. Regression analysis models are used to predict the value of one variable from one or more variables whose values can be predetermined. The line of regression gives the best estimates for the value of Y (dependent variable) for any specified value of X (independent variable). Here the variable which influences the value of another variable or used for prediction is called the independent variable and the value which is influenced and to be predicted is called the dependent variable. Hence we can forecast the total cost (dependent value) for a particular level of activity (independent variable) with the help of a regression equation.

4. Forecasts of costs and revenues using linear regression coefficients

In the regression analysis method, more than two points are considered for analysis, which reduces the chances of distortion in the results as are evident in the case of the high / low method.

In statistics, total cost is represented by the equation -

$y = mx + c$
Where,
y = total cost
m = variable cost per unit
x = number of units / volume of output
c = constant, signifying fixed costs

A line of regression can be drawn using this equation. From this equation, we can derive the following two equations in order to find out the value of m and c representing the variable cost per unit and the fixed cost respectively. These two equations are:

$$\Sigma y = nc + m\Sigma x$$

$$\Sigma xy = C\Sigma x + m\Sigma x^2$$

Where, x is the independent variable and y is the dependent variable

n = Number of observations
Σx = Sum of the values of the independent variable
Σy = Sum of the values of the dependent variable
Σx^2 = Sum of the squared values of the independent variable
Σxy = Sum of the products of x and y
m = Variable cost per unit
c = Fixed cost

By solving these two equations, we get the formula for calculating fixed and variable costs directly:

$$m = \frac{n\Sigma xy - \Sigma x\Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

$$c = \frac{\Sigma y - m\Sigma x}{n}$$

Example

The following are the sales and corresponding overhead costs of Peacock Ltd for a 5-year period. Identify the fixed and variable components of the costs.

Year	Sales Tshs'000	Overhead Costs Tshs'000
1	8,000	5,000
2	10,000	5,400
3	12,000	5,800
4	11,000	5,500
5	14,000	5,600
Total	55,000	27,300

Calculate the total cost for 15,000 units.

Answer

Let us consider x to be the independent variable representing sales and y to be a dependent variable representing overheads. We will use the above formulae to arrive at the fixed and variable components of the overheads.

$$m = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

$$c = \frac{\Sigma y - m\Sigma x}{n}$$

We need to calculate Σxy, Σx^2
Ignoring the figures up to '000 places,

Σxy = (8 x 5) + (10 x 5.4) + (12 x 5.8) + (11 x 5.5) + (14 x 5.6)
= 40 + 54 + 69.6 + 60.5 + 78.4
= 302.5

$\Sigma x^2 = 8^2+10^2+12^2+11^2+14^2$
= 625

$$m = \frac{[(5 \times 302.5) - (55 \times 27.3)]}{5 \times 625 - (55)^2}$$

$$m = \frac{1,512.5 - 1,501.5}{3,125 - 3,025}$$

m = 0.11 i.e. the variable cost per unit is \$0.11 per unit.

$$c = \frac{27.3 - (0.11 \times 55)}{5}$$

c = 4.25 i.e. \$4,250 i.e. the fixed cost is \$4,250.

Total cost for 15,000 units = \$0.11 x 15,000 + \$4,250 = \$5,900

3.7 Properties of linear regression

1. There are two regression equations; b_{yx} and b_{xy}. Both have the same sign and both cannot be greater than 1. For example, suppose two variables, x and y, are production units and man hours. The regression coefficient b_{xy} measures the extent of the relationship between production units and man hours whereas regression coefficient b_{yx} measures the extent of the relationship between man hours and production units. One of them can be greater than 1 but both cannot be greater than one at the same time. This is because correlation coefficient is the geometric mean of the regression coefficients i.e. $r = (+/-)\sqrt{b_{yx} \times b_{xy}}$ and r always lies between -1 and +1. In some cases, both of them can be smaller than 1.

2. Both regression coefficients will have the same sign as they depend upon the sign of Cov (x, y).

3. Regression coefficient is independent of origin but not of scale. It means if the values of x and y are too large, they can be minimised by subtracting some specific, assumed mean, but cannot be minimised by dividing any specific assumed mean.

4. Arithmetic mean of regression coefficients is greater than the correlation coefficient provided r > 0

5. If r=0, regression lines are perpendicular to each other. If r=+1 or -1, regression lines are tend to be parallel.

The following example will help you to understand most of the above listed properties:

Example

Given below is the data for Fantastic Ltd. Variable X denotes the number of years of experience of the employees and variable Y denotes the productivity index of the employees.

X	10	7	15	18	5
Y	65	60	72	75	57

Calculate two regression coefficients and estimate:

(i) The productivity index of an employee whose length of experience is 20 years
(ii) The number of years of experience of an employee whose productivity index is 100

X	Y	x = X - 11	y = Y - 65	x.x	y.y	x.y
10	65	-1	0	1	0	0
7	60	-4	-5	16	25	20
15	72	4	7	16	49	28
18	75	7	10	49	100	70
5	57	-6	-8	36	64	48
55	329	0	4	118	238	166

Here, $\Sigma x = 0$, $\Sigma y = 4$, $\Sigma x^2 = 118$, $\Sigma y^2 = 238$ and $\Sigma xy = 166$

Regression equation of Y on X: b_{yx}

$$= \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{5\,(166) - (0)\,(4)}{5\,(118) - (0)^2}$$

$$= \frac{830}{590}$$

= 1.406 = 1.41

Continued on the next page

Regression line of Y on X:

Y= a + bX
Y = 41.21 + 1.41X

Where,

$a = \overline{Y} - b_{yx}\overline{X}$

$= \frac{329}{5} - (1.41)\,\frac{55}{5}$

= 56.72 – 15.51

= 41.21

Let us calculate the productivity index of an employee whose length of experience is 20 years

Y = 41.21 + 1.41X

Y = 41.21 + 1.41 (20)

Y = 69.41

Regression equation of X on Y: b_{xy}

$$= \frac{nL\,xy - (L\,x)(L\,y)}{nL\,y^2 - (L\,y)^2}$$

$$= \frac{5\,(166) - (0)\,(4)}{5\,(238) - (4)^2}$$

$$= \frac{830}{1174}$$

= 0.706 = 0.71

Regression line of X on Y:

X = c + bY
X = -35.72 + 0.71Y

Where,

$c = \overline{X} - b_{xy}\overline{Y}$

$= \frac{55}{5} - (0.71)\,\frac{329}{5}$

= 11 – 46.72

= -35.72

Let us calculate the number of years of experience of an employee whose productivity index is 100

X = -35.72 + 0.71Y

X = -35.72 + 0.71 (100)

X = 35.28

Correlation coefficient

$r = (+/-)\sqrt{b_{yx} \times b_{xy}}$

$r = (+)\sqrt{1.406 \times 0.706}$

r = 0.9

Let us verify that Arithmetic mean of regression coefficients is greater than the correlation coefficient, provided r > 0

$$\frac{b_{yx} + b_{xy}}{2} = \frac{1.406 + 0.706}{2} = 1.05$$

Tip

In the examination, you may be provided with data relating to mean and standard deviation for both variables and the correlation coefficient. In such circumstances, regression coefficients can be derived as shown in the example below.

Example

The following data relates to Fortune Ltd. The management of the company wants to know that if a power cut lasts for 10 hours, what will be the estimated idle time of the employees.

	Mean	Standard deviation
Idle time of employees (hours)	40	10
Power cut in the factory (hours)	50	16

The correlation between the idle time of employees and the power cut in the factory is 0.5.

Let’s assume idle time hours as X and power cut hours as Y.

Regression coefficients and equation can be derived as follows:

$b_{yx} = r\frac{ay}{ax} = 0.5 \text{ x } \frac{16}{10} = 0.8$

Regression equation: Y = a + b X
Where, b = 0.8 and a = 50 – (0.8) (40) = 18

Y = a + b X
Y = 18 + 0.8 X

$b_{xy} = r\frac{ax}{ay} = 0.5 \text{ x } \frac{10}{16} = 0.32$

In accordance with the regression properties, alternatively, b_{xy} can be calculated as: $\frac{b_{yx}}{r^2}$

Regression equation: X = c + bY

Where, b = 0.3125 and c = 40 – (0.32) (50) = 24

X = c + bY
X = 24 + 0.32Y

By putting the value of Y (as 10 hours of power cut) in the regression equation:

X = c + bY
X = 24 + 0.32 (10)
X = 24.32 hours

Standard error of estimate

In a regression line, the smaller the standard error of the estimate, the more accurate are the predictions.

It is commonly used as SEE. The measure SEE depicts how well a linear regression model is performing. SEE compares the actual values in the dependent variable Y to the estimated values that would have resulted had Y followed the linear regression exactly.

This concept is similar to the concept of standard deviation. Standard error (SE) of an estimate is derived by calculating the mean of squares of deviations (between actual and estimated values) based on the regression analysis.

$$S_{yx} = \sqrt{\frac{\Sigma\left(Y - \hat{Y}\right)^2}{n-2}}$$

Similarly, S_{xy} can be derived

Where,

S_{yx} = Standard deviation of regression on Y values of X values

$\hat{Y}$ = actual value of variables

Y = estimated value of variables

n = number of observations

Example

Sigma Co used linear regression to predict Y (weight of certain persons) from X (height of those persons) in a certain population of 100 persons.

In this population, $\Sigma(Y - \hat{Y})^2$ is 50 and the correlation between X and Y is 0.5. What is the standard error of the estimate?

$$S_{yx} = \sqrt{\frac{\Sigma\left(Y - \hat{Y}\right)^2}{n-2}}$$

$$S_{yx} = \sqrt{\frac{50^2}{100-2}}$$

S_{yx} = 5.05

Example

A financial analyst of Gamma Ltd has developed a regression model that relates annual profit growth to company sales growth by the equation Y = 2.3 + 0.75X based on the experience of the last five years.

Year	Sales growth (X)	Actual profit growth ($\hat{Y}$)	Predicted profit growth (Y)	$(Y - \hat{Y})$	$(Y - \hat{Y})^2$
1	6.2	5.2	6.0	-0.8	0.64
2	4.8	2.7	2.5	0.2	0.04
3	0.9	1.1	0.7	0.4	0.16
4	3.2	2.0	2.4	-0.4	0.16
5	4.7	3.0	3.5	-0.5	0.25
	ΣX = 19.8	$\Sigma\hat{Y}$ = 14.0	ΣY = 15.1	$\Sigma(Y - \hat{Y})$ = -1.1	$\Sigma(Y - \hat{Y})^2$ = 1.3

$$SEE = \sqrt{\frac{\Sigma\left(Y - \hat{Y}\right)^2}{n-2}} = \sqrt{\frac{1.3}{5-2}} = 0.66$$

The lower the SEE is, the more accurate the predictions are.

We have already seen various examples in the previous Learning Outcomes to know how linear regression and coefficient analysis can be applied in real accounting and business scenarios. Here, we shall focus on some more complex examples to gain a deeper understanding of this topic!

Example

Lema Ltd has computed the correlation between the idle time of the employees and the productivity as -0.21.

While calculating the correlation coefficient between X and Y from a pair of 25 observations; the following summations have been taken:

X = 125
X^2 = 650
Y = 100
Y^2 = 460
Product of X and Y = 508

Later, it is detected that two pairs of observations were taken as (6, 14) and (8, 6) instead of (8, 12) and (6,8).

Derive the revised correlation coefficient.

Here, first we need to calculate the revised summations as follows:

ΣX = 125 + (8 + 6) – (6 + 8) = 125
ΣX^2 = 650 + (64 + 36) – (36 + 64) = 650
ΣY = 100 + (12 + 8) – (14 + 6) = 100
ΣY^2 = 460 + (144 + 64) – (196 + 36) = 436
ΣXY = 508 + (96 + 48) – (84 + 48) = 520

By applying the figures in the formula:

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

$$r = \frac{25\,(520) - (125)\,(100)}{\sqrt{\left(25\,(650) - (125)^2\right)\left(25\,(436) - (100)^2\right)}}$$

$$r = \frac{13000 - 12500}{\sqrt{(16250 - 15625)\,(10900 - 10000)}}$$

$$r = \frac{500}{\sqrt{(625)\,(900)}}$$

$$r = \frac{500}{750}$$

r = 0.67

Example

A regression line for sales forecasting derived on the basis of past experience of five years is:

Y = 343 + 0.2X.

Forecast the sales units for the second month of the sixth year.

Here,
X = 62 (as in the example it is given five years i.e. i.e. 60 months + second month of the sixth year i.e.2 months = 62 months

Sales forecast: Y = 343 + 0.2X
= 343 + 0.2(62)
= 343 + 12.4
= 355.4

Example

From the following statistics, forecast the value of Y for the third quarter of 20X2.

n = 100, ΣY = 1,525, ΣX = 2,536 and b = -2.36.

Hint: Consider X = 0; corresponds to the first quarter of 20X1.

$$a = \frac{\Sigma Y}{n} - (b)\frac{\Sigma X}{n} = 15.25 - (-2.36)\,(25.36) = 15.25 + 59.85 = 75.10$$

As mentioned, the value of X (for the third quarter of 20X2) will be 7

Y = a + bx
= 75.10 + (-2.36) (7)
= 75.10 – 16.52
= 58.58

Example

The line of regression of the mother’s height (X) on the daughter’s height (Y) in a survey of 50 persons is: 3Y – 5X + 180 = 0. The following information is available:

Average height of daughters is 44 meters and variance of a mother’s height is 0.5625th the variance of a daughter’s height.

Compute:

(a) Average height of mothers
(b) Correlation coefficient between mother’s height and daughter’s height

Regression equation X on Y is: 3Y – 5X + 180 = 0

5X = 3Y + 180
X = 0.6Y + 36
Hence, c = 36 and b_{xy} = 0.6

As we know, c = $\bar{X} - b\bar{Y}$
36 = $\bar{X}$ - (0.6) (44)
Hence, $\bar{X}$ = 62.5 (average height of mother)

Continued on the next page

As given in the question, the variance of a mother's height is 0.5625th the variance of a daughter's height. Hence, standard deviation of a mother's height is 0.75th the standard deviation of a daughter's height.

Symbolically,

$a_X = a_Y/0.75$

$$\frac{aX}{aY} = 0.75$$

Now, let us calculate the correlation coefficient

$$b_{xy} = r\frac{ax}{ay}$$

$0.6 = r\ (0.75)$

Hence, $r = 0.8$

The difference between these two statistical techniques, i.e. correlation and regression, can be summarised as follows:

Correlation	Regression
Correlation measures the strength of linear association between paired variables, say X and Y.	Regression measures the form of linear association that best predicts Y from the values of X and vice versa.
If variables X and Y are interchanged in the calculation of the correlation coefficient, the value of the correlation coefficient will remain the same.	Linear regression is not symmetric in terms of X and Y. This means interchanging X and Y will give a different regression model (i.e. X in terms of Y) against the original Y in terms of X.
Correlation coefficient value ranges from -1 to +1.	There are two regression coefficients; both cannot be greater than one.
Correlation coefficient is independent of both origin and scale.	Regression coefficient is independent of origin, but not of scale.

Test Yourself 8

Butterfly Plc manufactures talcum powder. The data on the monthly production of powder packs is given below.

The data in the table relates to the variable cost of manufacturing the powder packs. The fixed cost per year amounts to Tshs10,000,000.

Month	Production (packs)	Cost (Tshs'000)
January	5,000	55,000
February	6,000	64,000
March	7,000	73,000
April	8,000	82,000
May	9,000	91,000
June	10,000	100,000

Required:

Based on the information given, plot a line of best fit.

---

4. Calculate coefficient of determination.
   Interpret slope y intercept correlation coefficient and coefficient of determination.
   Conduct test for shape, and coefficient of correlation.

[Learning Outcomes d, e and f]

4.1 Coefficient of determination

Definition

The coefficient of determination is the square of the correlation coefficient and is expressed as r^2

The concept of "coefficient of determination" demonstrates the relationship between the coefficient of correlation and regression coefficients.

Mathematically the correlation coefficient is computed according to the above formula and the result is multiplied by itself to arrive at the coefficient of determination.

In the first question given above where, r = 0.997; the coefficient of determination can be calculated as:

$r^2 = (0.997)^2 = 0.994$.

The coefficient of determination, r^2, is useful because it gives the proportion of variance (fluctuation) of one variable that is predictable or which can be estimated from the other variable. It is a measure that enables us to determine how certain you can be in predicting outcomes from a given graph.

The coefficient of determination is the ratio of the explained variation to the total variation. The coefficient of determination lies between 0 and 1, both inclusive e.g. $0 \leq r^2 \leq 1$. Since 'r' ranges between -1 to +1, it denotes the strength of the linear association between x and y.

The coefficient of determination represents the percentage of data that is closest to the line of best fit.

Example

If r = 0.9, then $r^2 = 0.81$, which means that 81% of the total variation in y can be explained by the linear relationship between x and y. The other 19% of the total variation in y remains unexplained by the linear relationship.

The coefficient of determination is a measure of how well the regression line, that is the line joining the plotted points, represents the data. If the regression line passes exactly through every point on the scatter plot, it would be able to explain all of the variation.

Interpretation

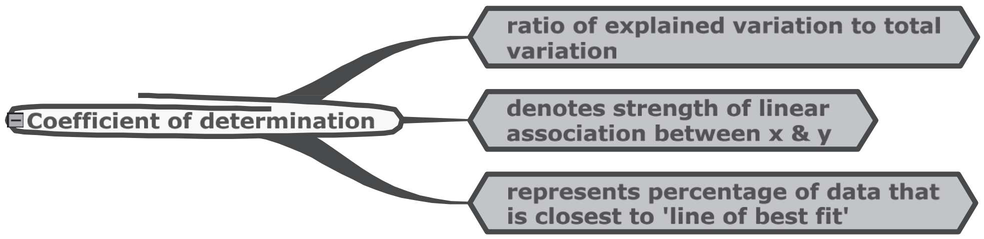
A correlation coefficient of 0.8 or more signifies strong correlation and a correlation coefficient of 0.5 or less signifies weak correlation.

However, unless the correlation coefficient is precisely or almost -1, +1 or 0, it cannot be interpreted well enough to establish the exact nature of relationship between two variables. Hence the coefficient of determination assists us to better understand the relation between variables.

Example

The correlation coefficient for two variables is 0.7. All we can figure out is that the two variables in question exhibit positive correlation. It is a partial positive correlation. However, if we square this coefficient, we get 0.49, that is, r^2 or the coefficient of determination. This signifies that 49% of variations can be explained by the key factor under consideration, and the rest can be explained by other factors.

SUMMARY



Test Yourself 9

The coefficient of determination:

A Is the square of the correlation coefficient
B Is the cube of the correlation coefficient
C Is arrived at by multiplying the correlation coefficient by 2
D None of the above

Test Yourself 10

If the coefficient of determination is 0.36, then correlation between variables is:

A 0.13
B 0
C 0.6
D None of the above

4.2 Decision on assigning y to the variable

Tip

a) It is often difficult for a student to decide which variable to assign y to, while solving regression analysis. In correlation, there is no impact on the answer if you assign y to the independent variable and x to the dependent variable or vice-versa, since it gives the same result. However, it does matter in regression. The forecast may show a wrong result if the denotation of the variables get changed.

b) This is because the regression line only minimises the sum of squares of the y-errors, and this is only equivalent to minimising x-errors in the case of perfect correlation, that is, when r = +1.

c) Variable y is the dependent variable and sometimes it is very clear which that is. In practice, if you wish to forecast a particular variable, that variable must be denoted by y.

4.3 Test for shape

This concept is discussed in the Study Guide 7, Hypothesis testing.

Answers to Test Yourself

Answer to TY 1

The correct option is D.

Exhibits all the above properties; coefficient is the factor that represents the correlation between variables. It depicts the extent to which the variables are related. It also underlines the influence of change of behaviour in one variable on another.

Answer to TY 2

The correct option is D.

Degrees of correlation between variables can be all of the above, that is, perfect, partial and absent. The correlation is perfect when the value of ‘r’ is -1 or +1. It is partial when the value of ‘r’ lies between -1 to +1. Correlation degree will be absent when ‘r’ = 0

Answer to TY 3

The correct option is A.

In a positive correlation the higher magnitudes of one variable correspond with the higher magnitudes of the other. The higher magnitudes of one variable, corresponding with lower magnitudes of the other denotes a negative correlation. A situation where lower magnitudes of one correspond with the higher ones of the other is again a situation of negative correlation.

Answer to TY 4

Even though the values of X and Y are not too large, we can use u and v to determine the correlation coefficient

X	Y	$u = \frac{X - 50}{5}$	v = Y - 4	u^2	v^2	uv
25	2	-5	-2	25	4	10
30	2	-4	-2	16	4	8
45	4	-1	0	1	0	0
50	5	0	1	0	1	0
55	6	1	2	1	4	2
60	7	2	3	4	9	6
65	8	3	4	9	16	12
L:X = 330	L:Y = 34	L:u = -4	L:v = 6	L: u^2 = 56	L: v^2 = 38	L:uv = 38

$$r = \frac{n L uv - L u L v}{\sqrt{\left(n L u^2 - (L u)^2\right)\left(n L v^2 - (L v)^2\right)}}$$

$$= \frac{(7 \times 38) - (-4 \times 6)}{\sqrt{((7 \times 56) - (16))((7 \times 38) - (36))}}$$

$$= \frac{266 + 24}{\sqrt{376 \times 230}}$$

$$= \frac{290}{\sqrt{86400}}$$

= 0.98 (strong positive correlation)

Answer to TY 5

The correct option is A.

In the regression analysis ‘n’ stands for the number of pairs of data for which the “line of best fit” is required to be plotted. �x^2 is not the same calculation as (�x) ^{2}; �x^2 is the sum of the values for x^2 as calculated and (�x) ^{2} is the ‘sum of x’ raised to 2.

Answer to TY 6

The correct option is A.

y = a + bx
= Tshs200,000 + Tshs2,500(350)
= Tshs200,000 + Tshs875,000
= Tshs1,075,000

Answer to TY 7

The correct option is D.

y = a + bx
Hence, 2,500,000 = a + b(200) and 4,900,000 = a + b(500)
By elimination,

4,900,000	= a + 500b
- 2,500,000	= a + 200b
2400,000	= 300b

Hence, b = 2,400,000/300 = 8,000
Putting the value of b into the equation
Tshs4,900,000 = a + (500 x Tshs8,000)
a = Tshs900,000

Answer to TY 8

Consider the month of January. Here, x = 5,000 and y = 55,000. Fixed cost is Tshs10,000,000, therefore a = 10,000, “b” can now be calculated as:

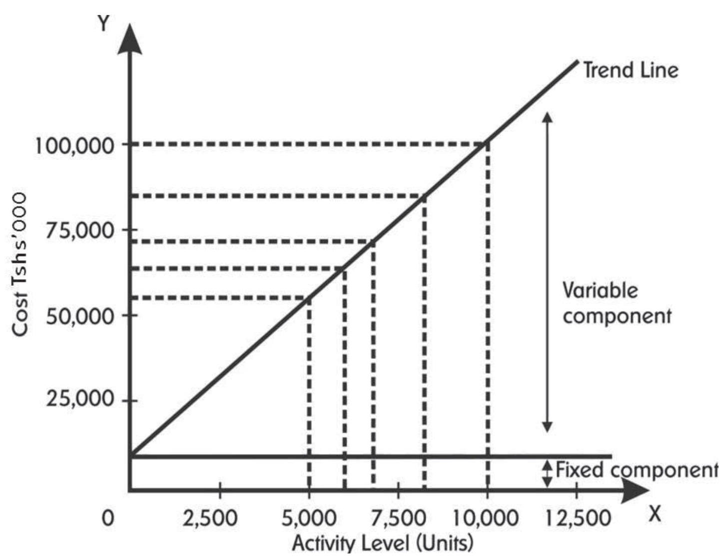
y = a + bx

55,000 = 10,000 + b (5,000)

5,000b = 45,000

b = 9

The line of best fit can be depicted on a graph as given below:



Taking the cost on Y axis and volume of units produced on X axis, we can plot the graph. By plotting all the points corresponding to the given activity levels we can draw a trend line by joining them. As we can see from the graph, it is a straight line, hence can be represented by a linear equation.

Answer to TY 9

The correct option is A.

It is the square of the correlation coefficient. All the other options are incorrect in relation to correlation coefficient.

Answer to TY 10

The correct option is C.

Correlation coefficient (r) = $\sqrt{\text{coefficient of determination}}$

$= \sqrt{0.36}$

$= 0.6$

Self Examination Questions

Question 1

Which of the following is a feasible value for the correlation coefficient?

A – 2.0
B – 1.2
C 0
D + 1.2

Question 2

The following information is provided for 3 pairs of data.
$\Sigma X = 6$, $\Sigma Y = 18$, $\Sigma XY = 32$, $\Sigma X^2 = 14$ and $\Sigma Y^2 = 116$

The coefficient of correlation will be:

A 1
B -1
C 0.08
D -0.08

Question 3

The coefficient of correlation is computed with the help of which of the following formulae?

A $r = \dfrac{n\Sigma XY - \Sigma X\Sigma Y}{\sqrt{(n(\Sigma X)^2 - \Sigma X^2)\,(n(\Sigma Y)^2 - \Sigma Y^2)}}$

B $r = \dfrac{n\Sigma XY - \Sigma X\Sigma Y}{\sqrt{(n\Sigma X^2 - (\Sigma X)^2)\,(n\Sigma Y^2 - (\Sigma Y)^2)}}$

C $r = \dfrac{n\Sigma X^2Y^2 - \Sigma X^2\Sigma Y^2}{\sqrt{(n\Sigma X^2 - (\Sigma X)^2)\,(n\Sigma Y^2 - (\Sigma Y)^2)}}$

D None of the above

Question 4

The correlation coefficient must have:

A Any desirable value
B A value between -1 and +1
C A value between -1 and 0
D All of the above

Question 5

Correlation is perfectly negative when:

A $r = +1$
B $r = 0$
C $r = -1$
D None of the above

Question 6

When the correlation coefficient lies between 0 and 0.5, the correlation is:

A High
B Strong
C Weak
D Low

Question 7

Correlation between two independent variables can be:

A Strong
B Moderate
C Weak
D None of the above

Question 8

Regression analysis is being used to find the line of best fit ($y = a + bx$) from ten pairs of data. The calculations have produced the following information:

$\Sigma x = 100$ $\Sigma y = 900$ $\Sigma xy = 25{,}000$ $\Sigma x^2 = 3{,}500$ $\Sigma y^2 = 29{,}500$

What is the value of 'a' in the equation for the line of best fit (to the nearest whole number)?

A 20
B 25
C 26
D 30

Question 9

Most of the documents of Robin Ltd were destroyed in a fire at the office. The following data of correlation analysis was available from the salvaged papers.

Variable X was assumed as advertising expenses and variable Y was assumed as sales volume.
Variance of advertising expenses: 9
Two regression equations: $8X – 10Y + 66 = 0$ and $40X – 18Y = 214$

Derive:
(a) The mean value of both variables, advertising expenses and sales volume
(b) Correlation between advertising expenses and sales volume
(c) Variance of Y, sales volume

Question 10

9 models participated in a beauty contest, which was judged by two panel members. The following are the ranks achieved by the models.

Models	M1	M2	M3	M4	M5	M6	M7	M8	M9
Judge A	9	5	1	4	2	7	8	6	3
Judge B	6	9	3	2	4	1	5	8	7

Required:

Plot a scatter diagram and interpret the correlation between the decisions of the two judges.

Answers to Self Examination Questions

Answer to SEQ 1

The correct option is C.

The value for ‘r’ cannot go beyond the range -1 to +1.

Answer to SEQ 2

The correct option is B.

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

$$= \frac{(3 \times 32) - (6 \times 18)}{\sqrt{((3 \times 14) - 36)((3 \times 116) - 324)}}$$

$$= \frac{96 - 108}{\sqrt{(42 - 36)(348 - 324)}}$$

$$= \frac{-12}{\sqrt{- -6 \times 24}}$$

$= -1$

Answer to SEQ 3

The correct option is B.

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

Answer to SEQ 4

The correct option is B.

The correlation coefficient must have a value between -1 and +1.

Answer to SEQ 5

The correct option is C.

r = -1. r = 0 implies no correlation and r = +1 implies perfectly positive correlation. So, these answers are incorrect.

Answer to SEQ 6

The correct option is C.

The closer the coefficient is to 1 or -1, the stronger the correlation is. Thus, when r is 0.5 or less it is said to be weak.

Answer to SEQ 7

The correct option is D.

There is no correlation between two independent variables.

Answer to SEQ 8

The correct option is C.

$$b = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{10(25,000) - (100)(900)}{10(3,500) - (100)^2}$$

$$= \frac{160,000}{25,000} = 6.40$$

$$a = \frac{\Sigma y}{n} - \frac{b\Sigma x}{n} = \frac{900}{10} - \frac{6.4(100)}{10} = 90 - 64 = 26$$

Answer to SEQ 9

(a)

$8X - 10Y + 66 = 0$
Hence,
$10Y = 8X + 66$
$Y = 0.8X + 6.6$
Hence,
$a = 6.6$ and $b_{yx} = 0.8$

As we know,

$a = \bar{Y} - b_{yx}\bar{X}$

$6.6 = \bar{Y} - 0.8\bar{X}$

$\bar{Y} = 6.6 + 0.8\bar{X}$

$40X - 18Y = 214$
Hence,
$40X = 18Y + 214$
$X = 0.45Y + 5.35$
Hence,
$c = 5.35$ and $b_{xy} = 0.45$

As we know,

$c = \bar{X} - b_{yx}\bar{Y}$

$5.35 = \bar{X} - 0.45\bar{Y}$

$\bar{X} = 5.35 + 0.45\bar{Y}$

By putting the value of $\bar{Y}$ in $\bar{X} = 5.35 + 0.45\bar{Y}$

$\bar{X} = 5.35 + 0.45\bar{Y}$

$\bar{X} = 5.35 + 0.45\,(6.6 + 0.8\bar{X})$

$\bar{X} = 5.35 + 2.97 + 0.36\bar{X}$

Hence,

$0.64\bar{X} = 8.32$

$\bar{X} = 13$

By putting the value of $\bar{X}$ in $\bar{X} = 5.35 + 0.45\bar{Y}$

$\bar{X} = 5.35 + 0.45\bar{Y}$

$13 = 5.35 + 0.45\bar{Y}$

$0.45\bar{Y} = 7.65$

$\bar{Y} = 17$

(b) Correlation regression:

$r = (+/-)\sqrt{b_{yx} \times b_{xy}}$

$= +\sqrt{0.8 \times 0.45}$

$= 0.6$

(c)

$b_{xy} = r\dfrac{ax}{ay}$

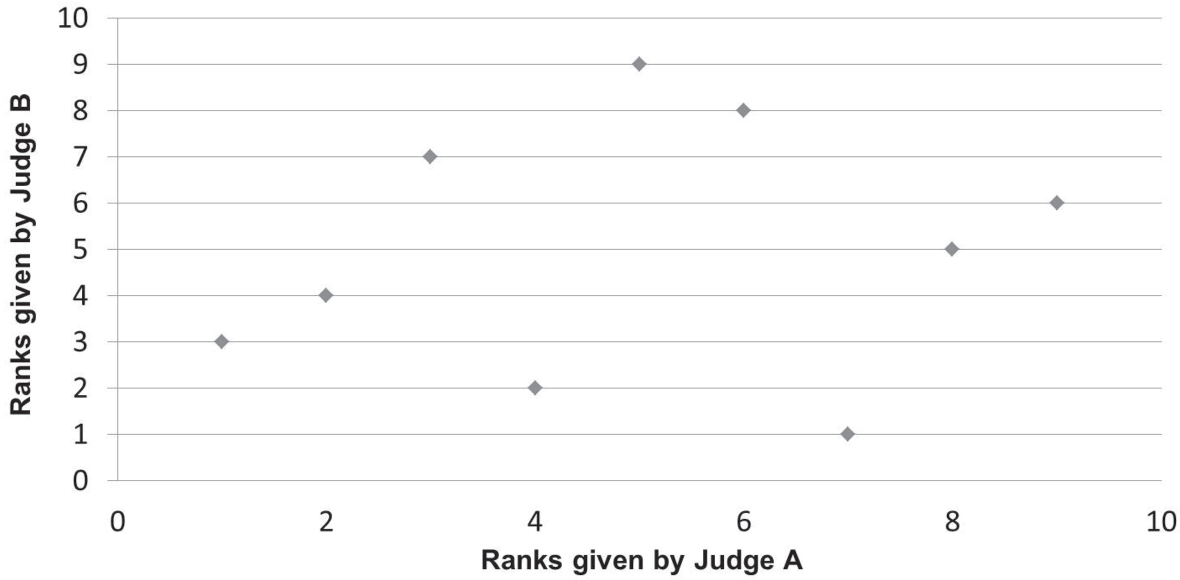
$0.45 = 0.6 \times \dfrac{\sqrt{9}}{aY}$

$aY = 4$

Hence, variance of Y = 16

Answer to SEQ 10

The scatter graph can be plotted as follows:



From the scatter diagram, it can be concluded that there is partial (very little) positive correlation between the judgements of both the judges.

(Note: if we calculate it by using the Spearman's rank correlation method, the answer would be 0.07.)

TIME SERIES

3

Get Through Intro

Statistical and mathematical tools not only help analyse past trends, but also assist in predicting future trends, thus helping immensely in planning for contingencies. These tools help management in developing a plan that can be as fool proof as possible.

Planning for any product does not end with planning for its production and use, but continues until the end of the product life cycle. This is an important concept in accounting and business, where the cost of a product over its complete life cycle is compared to the revenue generated from it in order to calculate the total costs and profits related to a product.

This section of the syllabus mainly aims at explaining the different techniques that help in the forecasting of sales and other such useful information such as costs and revenues associated with a product. Time series analysis is one such important technique that uses the study of a trend over a period of time in order to forecast and analyse product behaviour.

This Study Guide explores and discusses the types, uses and benefits of trend analysis. This knowledge will definitely help you study various products and provide accurate sales forecasts using these different techniques.

Learning Outcomes

a) Define time series and its characteristics.
b) Calculate trend using moving average method.
c) Calculate seasonal variations.
d) Forecasting future values.
e) Apply the concept of time series and forecasting in accounting and business situations.

1. Define time series and its characteristics.
[Learning Outcome a]

Forecasting is the basis of budgeting. The uncertainty in estimating the items of costs and revenues while preparing a budget can be dealt with by applying various forecasting techniques successfully. The forecasting methods can broadly be categorised into qualitative and quantitative techniques.

Definition

A time series is a series of data points i.e. figures or values recorded over time that indicates the trend of the data series.

Example

Monthly sales volume over the last five years; or sales during a particular season over the last ten years are examples of time series.

Time series analysis is one of the quantitative techniques of forecasting that aims to understand the trend of the data collected at regular intervals over a period of time, in order to project the trend for the future period. Under this method the past data is collected to understand its behaviour, and projected into the future to estimate the variable under consideration.

1.1 Characteristics of time series

1. A time series is a set of observations generated sequentially in time.
2. In a time series, data are collected to analyse the trend. The time periods are of equal length. For example, monthly, yearly, bi-annually, etc.
3. There are no missing values. For example, if the sales volume of the past 10 years is considered for analysis and the start year is 20X1, the end year should be 20Y0. One cannot start analysing sales trends starting with 20X1 and ending with 20Y2 by skipping any two years in between.
4. The time series is mostly applied in the areas of forecasting and determining the transfer function of a system.
5. Time series forecasting assumes that the basic forces underlying a time series (economic, political, social, or behavioural) are stable.
6. Usually, time series is helpful in forecasting:
    a) Economic planning
    b) Business planning
    c) Inventory control
    d) Production control

1.2 Components of time series

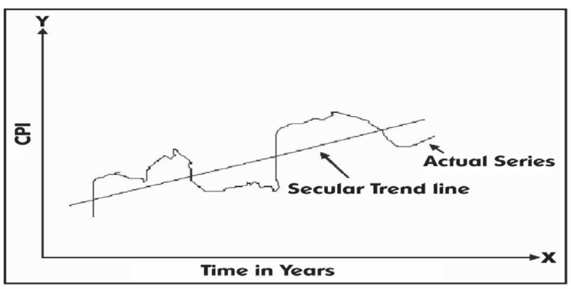
The variations observed in time series analysis (principles of time series) can be attributed to the following:

1. Secular trend

The secular trend is an accurate (without having any influence of cyclical fluctuations, seasonal variations and irregular variation) trend of data over a long period of time to analyse whether the value of the variable is decreasing, increasing or constant. From the actual time series, a trend line is plotted in order to understand the trend.

In the graph, we may notice that the variable is moving in a particular direction as indicated by the secular trend line.

Diagram 1: Secular trend of consumer price index



This gives an understanding of the behaviour of the variable and enables this behaviour to be projected into the future.

Example

An organisation has provided you with the following time series covering data over the last six years.

Year	Sales Tshs million	Profit Tshs million	Production Units
1	20	1.3	12,000
2	21.5	1.4	11,850
3	21	1.25	11,800
4	22	1.35	11,950
5	22.2	1.20	11,700
6	24	1.20	11,560

Let us find out the trend of changes in the value of the given variables over the period of time.

Sales revenue: From the given data, we can observe that there is an upward trend in sales revenue. Though there is a decrease in sales revenue between the years 2 and 3, the overall movement in sales revenue for the given period indicates a rising trend.

Profit: The time series has shown an increase from year 1 to the year 2, a decrease from year 3 to the year 4, and constant between year 5 and year 6. Hence there is no clear movement in the amount of profit over the given period of time.

Production volume: Observing the given time series, we can find that the production volume is decreasing every year except year 4, where production has increased. Hence, there is a downward trend in total production volume over the last 6 years.

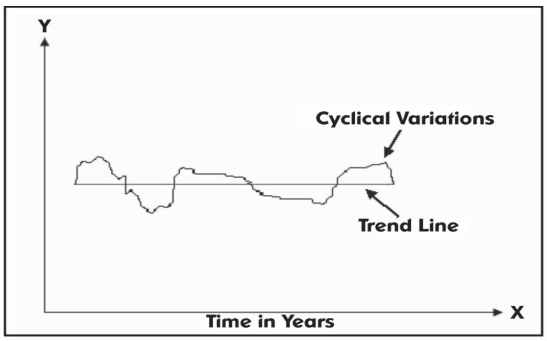
---

2. Cyclical fluctuations

Sometimes, the variations in the given data series might occur as a result of cyclical fluctuations. The most common reason for a cyclical fluctuation is the movement in the business cycle. Here too, the variable is analysed over a long period.

The future movement of the variable under consideration can be forecast taking into consideration the possible movements of the business cycle during the period under consideration (e.g. an anticipated boom in the economy would cause the trend line to move upwards and vice-versa).

Diagram 2: Cyclical variations

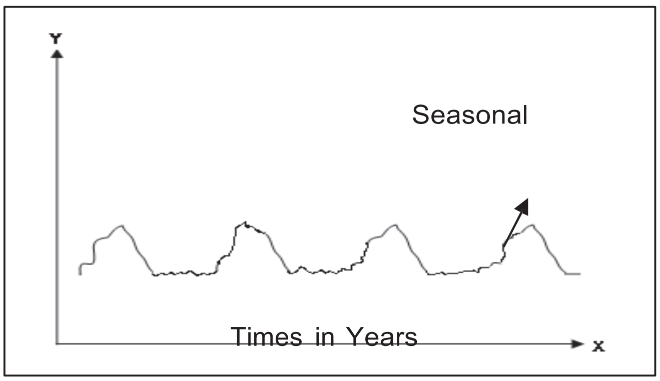


3. Seasonal variations

Seasonal variation is a component of time series which takes place in a short period of time. It indicates the regular and predictable fluctuations within a year or period of less than a year e.g. quarterly, monthly or weekly.

When a curve plotted, representing a time series shows a similar pattern during a certain period over successive years, it is indicative of seasonal variation.

Diagram 3: Seasonal variation

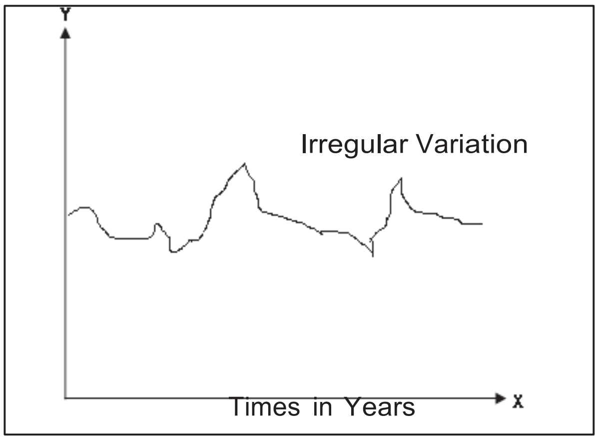


Example

If we observe the trend for the sale of winter garments over a number of years, it will be revealed that, in the winter season, the sales increase substantially in all the years under observation. This trend of raising sales during the winter season is attributable to seasonable variations. We can, therefore, predict that, in the coming years, the sale of winter garments will increase during the winter season.

4. Irregular variation

Diagram 4: Irregular variation (no trend)



Sometimes, the time series may not show any specific trend, and the variable may move inconsistently. It is difficult to pinpoint any reason for the erratic behaviour of the variable under consideration.

These kinds of variations are generally observed in cases where the time series are analysed for a few days or months and not over a long period.

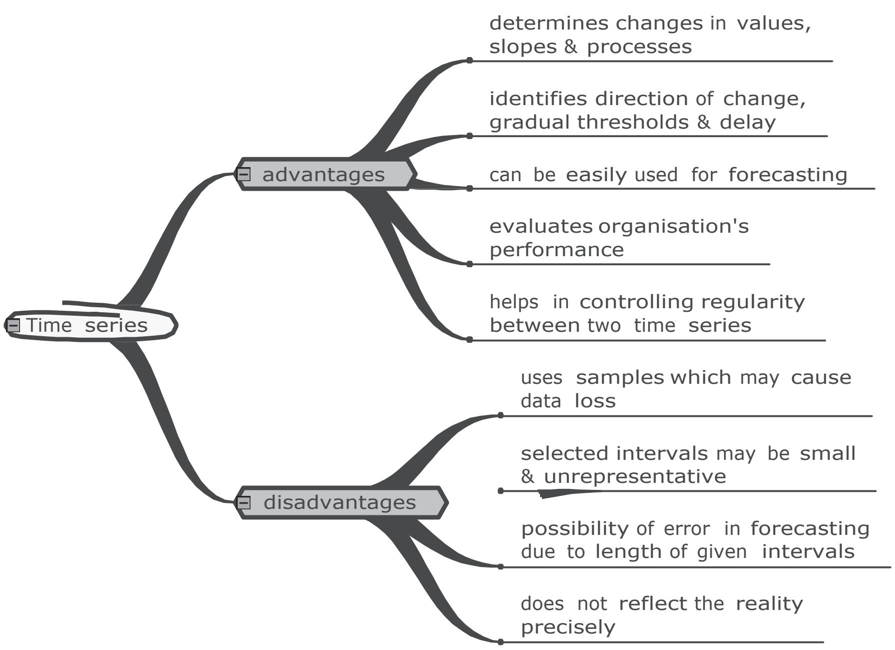
1.3 Advantages of time series analysis

1. Time series analysis has substantial advantages over other methods of analysis. It is a useful tool to determine changes in values, slopes and organisation processes over a period of time.
2. Time series analysis is important to identify the direction of change, gradual threshold and delay or change in outcome which might not be identified using other techniques.
3. Under time series, the study is based on readily available historical records. Hence it can be used for forecasting purposes more easily.
4. Time series analysis facilitates the evaluation of an organisation's performance. It compares the present value of a given variable with the trend over the past years, and thus indicates whether the performance of the firm has improved, weakened or remained constant over time.
5. Time series offers a set of procedures for analysing the cycle, synchronisation and relationship between two times series. This helps in controlling regularity between two time series.

1.4 Disadvantages of time series analysis

1. Time series selects a sample of time intervals which may cause data loss.
2. The time intervals selected for study purposes may be small and unrepresentative. This will not identify the causal relationship between the outcomes and time points accurately.
3. There is a possibility of error in forecasting due to the length of the time interval being analysed. This is because the Y values (dependent values) are appropriate, but the X values (independent values- time points) are approximate.
4. In time series analysis, neither the Y values nor the time points precisely reflect the reality.

SUMMARY



Example

Which of the following statement is/are correct?

(i) Time series analysis is a quantitative method of forecasting.
(ii) A seasonal variation is a component of time series that indicates the regular and predictable movements in the value of a given variable over a short period of time.
(iii) Time series analysis identifies the changes and direction of change in values of variables over a period of time.

A (i)
B (i) and (ii)
C (ii) and (iii)
D All of the above

Test Yourself 2

Which of the following best defines a time series?

A A series of events taking place within a specified time frame
B A given period of time for which data is collected and analysed
C An infinite time period for information is to be collated and evaluated
D All of the above

2. Calculate trend using the moving average method.
Calculate seasonal variations.
Forecasting future values.

[Learning Outcomes b, c and d]

2.1 Calculation of moving averages

One of the methods of finding a trend is the moving average method. Moving averages are widely used in time series analysis for forecasting purposes. It is an average of the different outcomes of a given number of periods. The moving average method redistributes the results obtained over a wider period of time to remove the effect of random and large variations from time series data. Moving averages remove the effects of seasonal variations and smooth the data set for further analysis.

For a given time series, the moving total is calculated for a fixed number of periods (e.g. three years or three quarters etc.). From the moving total, the average for that period is calculated.

Let us understand the calculation of moving averages with the help of the following example.

Example

The following are the sales figures for the last five years. Calculate the moving average for three years.

Years	Sales (in units)
20X4	200
20X5	250
20X6	225
20X7	260
20X8	280

Answer

Years	Sales (in units)	Moving Total	Moving average
20X4	200		
20X5	250	675 (200 + 250 + 225)	225 (675/3)
20X6	225	735 (250 + 225 + 260)	245 (735/3)
20X7	260	765 (225 + 260 + 280)	255 (765/3)
20X8	280		

When the set of data considered in a time series is odd, the moving averages are already centred as shown in the above example, but when the set of data considered is even, the moving averages do not become centred. In this case, we need to calculate the centred moving average.

When calculating a moving average, placing the average in the middle of the time period makes sense. If we average an even number of terms, we need to smooth out the value. To do this, the centred moving average is calculated by taking the mean of the two moving averages.

Example

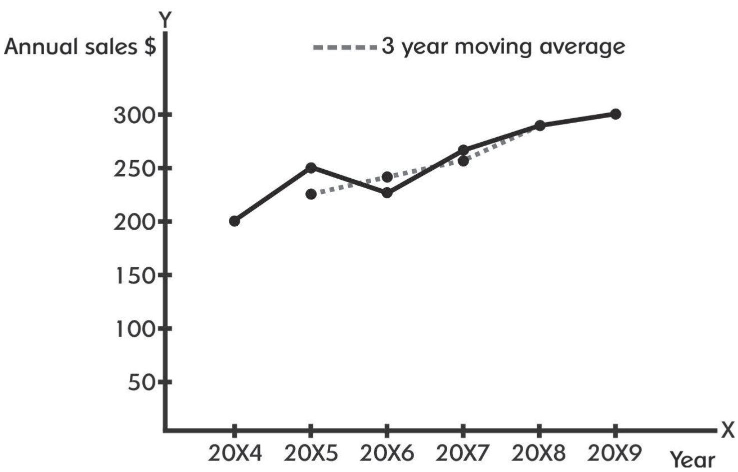
In the above example, if we add one more year, and the corresponding sales are 300, the centred moving average would be calculated as follows:

Years	Sales (in units)	Moving Total	Moving average	Centred moving average
20X4	200			
20X5	250	675	225	
				235 [(225 + 245)/2]
20X6	225	735	245	
				250 [(245 + 255)/2]
20X7	260	765	255	
				267.5 [(255 + 280) /2]
20X8	280	840	280	
20X9	300			

The graphical presentation of moving averages is explained below with the help of the following example.

Example

Continuing with the previous example



2.2 Seasonal variations (additive and multiplicative) to make budget forecasts

There are two models of time series analysis: the additive model and the multiplicative model, used to estimate the seasonal variations.

1. Additive model

In order to forecast the future by using time series analysis, we consider various factors influencing the variable over the time period under consideration. They could include secular trend, cyclical variations, seasonal variations and irregular fluctuations.

Under the additive model, it is assumed that the factors of the series i.e. seasonal variations, secular trend, cyclical variations and irregular fluctuations are independent of each other and do not have any impact on each other.

Example

The trend of sales of shoes may be attributable to factors such as the trend of the growth of the business, the trend of a boom in the business cycle, the growth of sales of shoes during festive seasons and other unknown factors causing erratic behaviour of the overall trend of sales. However, these factors act independently on the trend and do not influence each other.

Therefore under this method, time series is predicted using the formula below:

$\boxed{y = T + S + C + I}$

Where,
y = the actual time series
T = the trend series
S = the seasonal component
C = the cyclical component
I = the irregular component

(a) Finding out the seasonal trend from the additive model

The effect of each phase of cyclical variation is felt through a long period of time, about 10 to 12 years, whereas the future is generally forecasted for a shorter period which may be a quarter, six months or at the most a year. Therefore, the effect of cyclical variation on the variable may not be significant.

Also, in time series analysis, generally, a moving average is used, in which the earliest year is eliminated and a new year is added. This takes care of the impact of cyclical variations on the time series analysis.
Therefore, the time series formula may be written, ignoring the effect of the cyclical variation as: y = T + S + I.

Once the moving or centred moving averages are calculated, variance is calculated and, on the basis of these variances, adjustments are made to reduce the effect of seasonal variations.

It is assumed that irregular fluctuations will have a negligible effect on the trend, and hence can be ignored. Therefore, actual time series can only be ascertained by considering secular trend and seasonal variations.
Seasonal variations can be calculated as S = y – T.

Example

The following is the quarterly data for the last three years' sales (Tshs million) of Ramco Plc. Find out the seasonal components of the trend using the additive model.

Year	Actual			
20X6	I	II	III	IV
	420	440	500	620
20X7	I	II	III	IV
	450	460	540	660
20X8	I	II	III	IV
	500	490	590	690

Answer

Years	Quarters	Actual Y	Moving Average	Centred Moving Average T	Seasonal Variations (Y – T)
20X6	I	420			
	II	440			
			495		
	III	500		498.75	1.25
			502.5		
	IV	620		505	115
			507.5		

Continued on the next page

20X7	I	450		512.5	(62.5)
			517.5		
	II	460		522.5	(62.5)
			527.5		
	III	540		533.75	6.25
			540		
	IV	660		543.75	116.25
			547.5		
20X8	I	500		553.75	(53.75)
			560		
	II	490		563.75	(73.75)
			567.5		
	III	590			
	IV	690			

Notes:

1. Calculation of moving averages :

(420 + 440 + 500 + 620)/4 = 495
(440 + 500 + 620 +450)/4 = 502.5

Other moving averages are calculated in the same manner.

2. Calculation of centred moving averages:

(495 + 502.5)/ 2 = 498.75
(502.5 + 507.5)/2 = 505

Subsequent centred moving averages are calculated in the same manner.

(b) Predicting future value of time series

To project the actual trend into the future, the seasonal variations need to be identified. The seasonal variations over the periods will not be the same, but the average of the variations for the preceding similar periods (e.g. quarter-wise) can be taken as representative figures.

Example

Continuing with the above example of Ramco PLC

Year	Quarters				Total
	I	II	III	IV	
20X6			1.25	115	
20X7	(62.5)	(62.5)	6.25	116.25	
20X8	(53.75)	(73.75)			
Averages	(58.125)	(68.125)	3.75	115.625	(6.875)

(c) Projecting the trend

The sum of the variations due to seasonal factors around the basic trend line should be zero in ideal conditions. Therefore, we need to adjust -6.875 in total i.e. -6.875/4 = -1.72 per quarter to have the sum of the quarterly variances as zero.

Example

Continuing the above example of Ramco Plc

					Total
Averages	(58.125)	(68.125)	3.75	115.625	(6.875)
Adjustment	1.72	1.72	1.72	1.72	6.875
Trend of seasonal variations	(56.405)	(66.405)	5.47	117.345	0

The moving average trend for the projected period is identified by averaging out the increase in the secular trend during the period of observation.

$$\text{Average increase in trend} = \frac{\text{Last centred moving average - first centred moving average}}{\text{Number of time periods - 1}}$$

It comes out to be (563.75 - 498.75)/(8-1) = 65/7 = 9.2857.

The actual trend may be calculated as follows:

Quarters	20X8 III	20X8 IV	20X9 I	20X9 II
Base trend (central moving average for II quarter, 20X8)	563.75	563.75	563.75	563.75
Add: Quarterly increasing trend	9.2875	18.575 (9.2875 x 2)	27.863 (9.2875 x 3)	37.15 (9.2875 x 4)
Actual trend	573.0375	582.325	591.613	600.9
Quarterly projected sales	Tshs573,037.50	Tshs582,325	Tshs591,613	Tshs600,900

2. Multiplicative model

Under this method, it is assumed that the components of time series have an impact on each other, which therefore multiplies the effect.

The formula used under this method is:

$\boxed{y = T \times S \times I \times C}$

Under this model, seasonal variations can be calculated as S = y/T, ignoring the irregular and cyclical variations.

Example

Forecast the sales of Fortune Ltd from the given quarterly data of the last three years, assuming the multiplicative model

Years	20X4				20X5				20X6			
Quarters	I	II	III	IV	I	II	III	IV	I	II	III	IV
Sales (in units in '000)	280	360	350	310	320	400	370	350	340	420	390	360

Answer

Years	Quarters	Actual (a)	Moving Average (b)	Centred moving Average (c)	Seasonal variations percentage (d)=[(a)/(c)] x 100
20X4	I	280			
	II	360			
			325		
	III	350		330	106
			335		
	IV	310		340	91
			345		
20X5	I II	320		347.5	92
			350		
	III	400		355	113
			360		
	IV	370		362.5	102
			365		
		350		367.5	95
			370		
20X6	I II	340		372.5	91
			375		
	III	420		376.25	112
			377.5		
	IV	390			
		360			

Year	Quarters				Total
	I	II	III	IV	
20X4			106	91	
20X5	92	113	102	95	
20X6	91	112			
Averages	91.5	112.5	104	93	401

For all the quarters, the modified mean percentage should come to 400 assuming 100 for each quarter. But the sum is 401; therefore, an adjustment of 0.25 per quarter is needed.

	Quarters				Total
	I	II	III	IV	
Averages	91.5	112.5	104	93	401
Adjustment	(0.25)	(0.25)	(0.25)	(0.25)	(1)
Trend of seasonal variations	91.25	112.25	103.75	92.75	400

The average percentage increase in the trend of centred moving average = (376.25 - 330.00)/7
= 6.607

Forecast centred moving average trend = 376.25 x 106.607%
= 401.11

Forecast sales for quarter I of year 20X7 = 401,110 x 91.25% = 36,6012.875
= 366,013 units

2.3 Use of regression analysis to determine trends and forecasting future values

As in time series analysis, a trend is represented by a straight line, the following equations are used to predict the future:

$\Sigma y = nc + m\Sigma x$

$\Sigma xy = c\Sigma x + m\Sigma x^2$

In time series analysis, the independent variable is time. In the given equations, x is the independent variable representing time, and y represents the variable that is dependent on time. The values of y can be obtained for every value of x by identifying and replacing the values of m and c in the equation of the straight line. Here m is the variable element (co-efficient of x) and c is the fixed element.

Therefore, $y = mx + c$

Let us take an example to understand how future outcomes can be predicted using the regression equations under time series analysis.

Example

The following is the sales volume data of Compact Ltd, which has been selling personal computers for the last 5 years. With reference to this data, predict the sales volume for the sixth year.

Period	Sales volume ('000 units)
20X2	30
20X3	32
20X4	34
20X5	36
20X6	38

Answer

In order to predict the sales volume for the next year, we will use the following formulae -

$\Sigma y = nc + m\Sigma x$

$\Sigma xy = c\Sigma x + m\Sigma x^2$

As we know, in time series x, the independent variable represents time and the dependent variable y, represents sales volume. For the convenience of solving the problem, we replace x by X where X = x – 20X1

Therefore,

X = 1, 2, 3, 4, 5.
n = 5
$\Sigma X = 1 + 2 + 3 + 4 + 5 = 15$
$\Sigma y = 30 + 32 + 34 + 36 + 38 = 170$

Σxy and Σx^2 are calculated below -

	X	y	Xy	X^2
	1	30	30	1
	2	32	64	4
	3	34	102	9
	4	36	144	16
	5	38	190	25
Total	15	170	530	55

Continued on the next page

Put these values in the relevant places of the formulae:

170 = 5c + 15m (equation 1)

530 = 15c + 55m (equation 2)

Multiply equation (1) by 3 to get:

510 = 15c + 45m (equation 3)
530 = 15c + 55m

Subtract equation 3 from equation 2 to get:

20 = 10m
Therefore, m = 2

Put the value of m in equation (1) to get the value of c i.e. c = 28.

Put the values of m and c in the straight line equation i.e. y = mx + c to calculate 'y' i.e. the volume of sale with respect to a given year.

Therefore, for the sixth year,

y = (2 x 6) + 28 (Being the calculation of y for 20X6, X would be 6)
y = 40

For 20X7 the sales volume is expected to be 40,000 units.

Test Yourself 3

Refer to the example of Ramco Plc (taking the last two years data) and find out the seasonal variations using the multiplicative model.

Test Yourself 4

Based on linear regression analysis, the quarterly trend in sales units for SuperDol Ltd may be represented by the equation: y = 2,500 + 80x

Where,

y = forecast sales trend in cases per quarter
x = the quarterly numbers, where the first quarter of the year 20X0 = 1, the second quarter of the year 20X0 = 2, etc.

The average seasonal variation in sales follows an additive model with the following quarterly variations.

Quarter	1	2	3	4
Seasonal Variation (Cases)	+50	–40	+50	-60

Forecast sales for the last quarter of 20X4.

A 2,500 units
B 2,440 units
C 4,100 units
D 4,040 units

---

3. Apply the concept of time series and forecasting in accounting and business situations. [Learning Outcome e]

The time series technique is widely used in accounting and business situations. It is commonly used by financial managers to forecast financial situations.

Businesses involved in finance related matters, such as banks, financial institutions, etc, depend greatly on forecast of future events like change in rates, fluctuations in the amount borrowed and lent, etc.

Major financial decisions include lending and borrowing money from various sources. All these rely on forecasts of future cash flow and future expected returns.

Example

A financial institution agrees to lend an organisation a certain amount of loan. At that time, it was assumed that the organisation is expected to return the loan amount in pre-determined instalments to the financial institution.

Time series is a technique that financial managers use to predict the likely future values of financial variables such as revenues (in-flows), expenses (out-flows), and cash balances.

Time series is also useful in sales and production forecasting.

Example

Tamari and Zakayo Ltd uses an additive time series analysis to forecast sales volume. The trend in sales and forecast seasonal variations for 20X2 are given below.

Quarter	January to March	April to June	July to September	October to December
Trend (sales units)	24,000	26,000	28,000	30,000
Seasonal variation (sales units)	(3,000)	4,000	6,000	(7,000)

The sales trend figures for the first two quarters of 20X3 are estimated at 32,000 and 34,000 units respectively. Quarterly seasonal variations are expected to be the same as the year 20X2.

Sales forecasting for 20X3 using the additive model

Quarter	January to March	April to June	July to September	October to December
Trend (sales units)	24,000	26,000	28,000	30,000
Seasonal variation (sales units)	(3,000)	4,000	6,000	(7,000)
Forecast sales volume	21,000	30,000	34,000	23,000

The application of time series in not just limited to sales, production and forecast only. It is also extensively applied in predicting business cycles, product life cycles, etc.

(i) In economics, it is said that any product passes through four stages; introduction, growth, stability and decline. The trend analysis here helps an organisation to estimate the period of each stage. Furthermore, it also helps to predict the sales and profitability at each stage.

(ii) Stock markets often use trend analysis to forecast the fluctuations in the share prices.

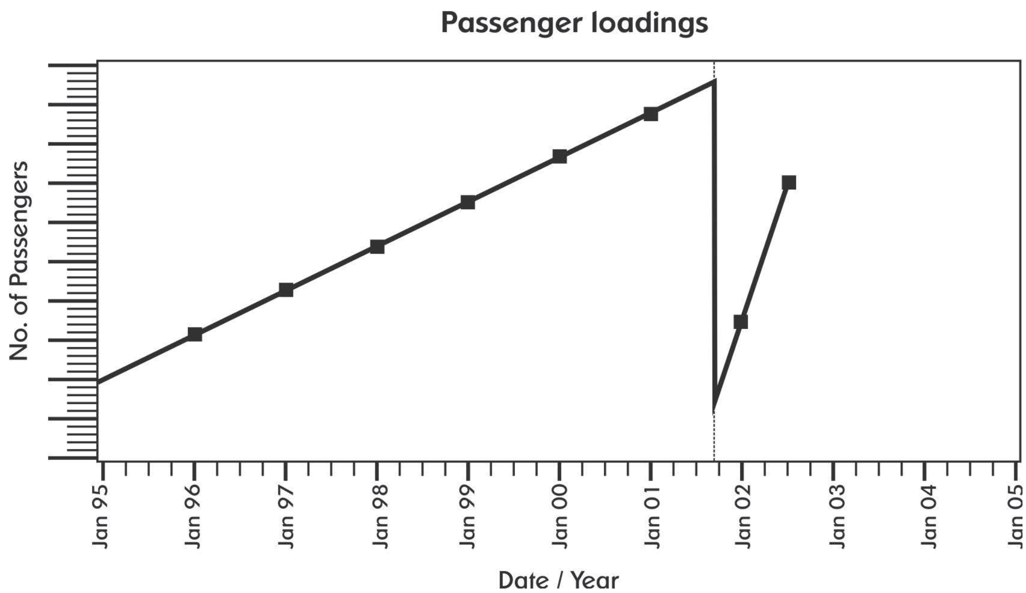
(iii) Advanced time series model also aids in forecasting demography and economic growth. For demography study, annual data of mortality rate, fertility rate, birth rate and death rate are collected. For economic growth predictions, annual data of GDP, inflation rate, employment rate etc. are collected.

The following case study illustrates the practical usage of time series.

Case Study

Most of us are aware of the world trade centre attacks of 9/11. The attacks obviously had a great impact on the passenger loadings in the air travel industry. Because of the closure of airports for some time and the subsequent public reaction, the average passenger loadings went down substantially. Over a period of time, public confidence was regained to some extent, and the passenger loadings went up.

All this can be seen from the following diagram:



The above diagram illustrates the use of time series in the study of trends and their use in future predictions. The above case is a historical event. The trend derived using time series analysis is applied by various airlines to predict future issues related to the air travel industry.

We can conclude that time series analysis is an important tool for the management in strategic decision making, especially under uncertainty.

Test Yourself 5

A bank has analysed the customer data relating to customers who have taken long-term loans and are irregular in the payments of installments and interest. By using three years' moving average of the year 20X1, seasonal variations have been derived as follows:

	Tshs'000
Quarter 1	+270
Quarter 2	-340
Quarter 3	+480
Quarter 4	-630

The loan disbursement during the year 20X4 is anticipated as Tshs5,000,000.

Required:

If the outstanding instalment and interest amounts are presented as: y = 200 + 1.6X (amounts in Tshs million); forecast the outstanding loan amount for the fourth quarter.

Answers to Test Yourself

Answer to TY 1

The correct option is D.

- Time series analysis is one of the quantitative techniques of forecasting that aims to understand the trend of the data collected at regular intervals over a period of time, to make forecasts.
- Seasonal variation is a component of time series, which indicates the regular and predictable fluctuations within a year or period of less than a year, quarterly, monthly or weekly.
- Time series analysis has substantial advantages over other methods of analysis. Time series analysis is a useful tool to determine changes in values and the slope and direction of changes over a period of time.

Answer to TY 2

The correct option is B.

Time series means a given period of time for which data is collected and analysed. Moreover, it is a bounded series and not an infinite time period.

Answer to TY 3

Years	Quarters	Actual Y	Moving T	Seasonal Variation % Y/T
20X7	I	450	512.5	87.80
	II	460	522.5	88.04
	III	540	533.75	101.17
	IV	660	543.75	121.38
20X8	I	500	553.75	90.29
	II	490	563.75	86.92
	III	690		

Answer to TY 4

The correct option is D.

Trend line: y = 2,500 + 80x
Here, x = last quarter of 20X4 i.e. 20

Y = 2,500 + 80 (20)
= 2,500 + 1,600
= 4,100

Effect of seasonal variation is -60 in the last quarter

Hence,
Sales forecast for the last quarter of 20X4 = 4,100 - 60 = 4,040

Answer to TY 5

Trend line: y = 200 + 1.6X
Where, x = quarter 4 of 20X4 i.e. 16
Y = 200 + 1.6 (16)
Y = 200 + 9.6
Y = 9.6 i.e Tshs9,600,000

Effect of seasonal variation is -630,000 in the last quarter

Hence,
Outstanding loan amount forecast for the last quarter of 20X4
= Tshs9,600,000 – Tshs630,000
= Tshs8,970,000

Self Examination Questions

Question 1

All of the following components could be included in a time series based sales forecast, except:

A Random fluctuations
B Seasonal variations
C Trends
D Cyclical variations

Question 2

Manyara region exports wheat every year in January, April, July and October to Dar-es-Salaam

Given below is the data regarding the value of wheat exported during the last four years on a quarterly basis in Tshs million.

(Wheat in Tshs million)				
Year	January	April	July	October
20X5	150	250	250	150
20X6	350	250	450	350
20X7	650	750	850	850
20X8	450	550	550	650

Required:

For this data, determine the seasonal variations using the multiplicative model of trend series.

Question 3

JD Pharmacy ltd has provided the annual sales data as below.

Year	1	2	3	4	5	6	7	8	9	10
Sales units (million)	40	42	46	44	50	48	54	52	56	60

Required:

Determine the seasonal variations using the addictive model, assuming a four yearly cycle.

Question 4

Define the terms ‘trend’ and ‘seasonal variation’ in the context of a time series analysis of sales.

Answers to Self Examination Questions

Answer to SEQ 1

The correct option is A.

Random fluctuation/s could not be included in a time series based sales forecast.

Answer to SEQ 2

Year	Month	Actual $ million	Four quarter moving total	Four quarter moving average	Centred moving average (CMA)	Percentage of actual to CMA
20X5	January	150				
	April	250				
			800	200		
	July	250			225	111%
			1,000	250		
	October	150			250	60%
			1,000	250		
20X6	January	350			275	127%
			1,200	300		
	April	250			325	77%
			1,400	350		
	July	450			387.5	116%
			1,700	425		
	October	350			487.5	72%
			2,200	550		
20X7	January	650			600	108%
			2,600	650		
	April	750			712.5	105%
			3,100	775		
	July	850			750	113%
			2,900	725		
	October	850			700	121%
			2,700	675		
20X8	January	450			637.5	71%
			2,400	600		
	April	550			575	96%
			2,200	550		
	July	550				
	October	650				

Answer to SEQ 3

Year	Sales units (Tshs million) Y	4 year Moving average	Centred moving average T	Seasonal variations Y – T
1	40			
2	42			
		43.00		
3	46		44.25	1.75
		45.50		
4	44		46.25	(2.25)
		47.00		
5	50		48.00	2.00
		49.00		
6	48		50.00	(2.00)
		51.00		
7	54		51.75	2.25
		52.50		
8	52		54.0	(2.00)
		55.50		
9	56			
10	60			

Answer to SEQ 4

In the context of time series analysis of sales:

a) Trend refers to the overall direction in which a graph of sales goes over a long interval of time.

b) Seasonal variations are the identical patterns of sales which occur in a short period of time. It indicates the regular and predictable fluctuations within a year or period of less than a year e.g. quarterly, monthly or weekly. For example, each year, many retail businesses experience an increase in sales volume before any festive season, and then there is a fall in sales after the festive season.

PROBABILITY DISTRIBUTION

4

Get Through Intro

The theory of probability has its origin in the games of chance related to gambling such as tossing a coin, throwing a die, drawing cards from a pack etc. A gambler's dispute in 1654 led to the creation of a mathematical theory of probability by two famous French mathematicians, Blaise Pascal and Pierre de Fermat.

Probability, which means likelihood or possibility, is a measure of how likely it is that something will happen. Probabilities are given a value between 0 (0% chance or will not happen) and 1 (100% chance or will happen).

Nowadays, the theory of probability is applied in everyday life in risk assessment, in trading on financial markets and in the solutions to social, economic and business problems.

Learning Outcomes

a) Explain in brief the concept of probability.
b) Calculate probability of events using binomial, Poisson and normal distribution.
c) Apply probability concept in accounting and business situations.

1. Explain in brief the concept of probability.

[Learning Outcome a]

Basic terms associated with the study of probability are:

1. Experiment: an experiment is any action / process resulting in one or more outcomes. The outcomes of an experiment can either be predictable with surety or random.

2. Deterministic experiment: is one for which outcome is unique and can be predicted in advance. For example, if an individual has availed of a loan of Tshs100 million from a bank at 15% rate of interest, the interest and principal outflow are easily determinable.

3. Random experiment: is an experiment (i) which can be repeated under identical conditions and (ii) which has two or more mutually exclusive outcomes which are known in advance - but it is not possible to predict which of them will result in any particular performance of the experiment. For example, tossing of a coin till head occurs or drawing out a card from a pack of playing cards, etc. are all random experiments.

4. Sample space: the set of all possible outcomes of a random experiment is called sample space. It is usually denoted by the letter 'S'.

5. Events: the outcomes of a random experiment are called events.

Definition

Probability: let S be an equiprobable sample space consisting of n distinct outcomes. If A is any event defined on S which consists of m ($ n) elements, then the probability of event A is given as:

$$P(A) = \frac{\text{Number of events favourable to A}}{\text{Total number of outcomes in the sample space}} = \frac{m}{n}$$

Note:

Since 0 $ m $ n, 0 $ P(A) $ 1, therefore the probability of any event lies between 0 and 1.
P(A') = 1- P(A) where A' is the complement of A.

Tip

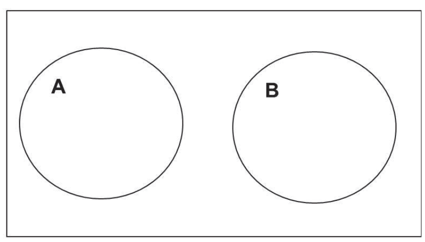
Probability can be expressed in many different ways. For example, as a proportion: ½; as a percentage 50% or as a decimal 0.50.

Example

The probability of getting a diamond or an ace not of diamond from a pack of 52 cards can be calculated as follows:

A pack of 52 cards containing 13 cards of each shape. Each set of 13 cards of a particular shape have one ace. So the number of total events is 52 and those favourable to the outcome are 13 diamonds plus 3 aces of other shapes = 16. Therefore, P(A) = n(A)/ n(S) = 16/ 52 = 4/13.

6. Mutually exclusive events: two events A and B defined on a sample space S are said to be mutually exclusive it they cannot occur simultaneously, i.e. there is no common element in A and B. Mutually exclusive events are also known as disjoint events. For example, in an experiment of drawing a card from a pack of cards, if an event of drawing a club card is denoted as A and an event of drawing a red card is denoted as B, A ∩ B is a null set.



7. Exhaustive events: the events A and B are said to form an exhaustive set when one of them must necessarily occur include if A and B are two events associated with a random experiment then when they are exhaustive we have AUB = S union of events gives the sample space

8. Equally probable, equally likely or mutually symmetric events: the events of an experiment are said to be equally probable when no event is expected to occur more frequently as compared to other events. For example, when a coin is tossed the outcomes head and tail are examples of such events.

9. Complementary events: the event 'A occurs' and the event 'A does not occur' are called complementary events to each other. The event 'A does not occur' is denoted by A¢ or A' or A^c. The event and its complements are mutually exclusive. For example, while tossing a coin, the event of getting a head is called A and the event of getting a tail is A'.

10. Dependent events: two events, A and B, are dependent if the outcome or occurrence of the first affects the outcome or occurrence of the second, so that the probability is changed.

Example

Two cards are chosen at random from a deck of 52 cards without replacement. What is the probability of choosing two jacks?

Let the probability of choosing the first jack be denoted as P (A)

$$P(A) = \frac{m}{n} = \frac{{}^{4}C_1}{{}^{52}C_1} = \frac{4}{52} = \frac{1}{13}$$

Let the probability of choosing the second jack be denoted as P (B)

$$P(B) = \frac{m}{n} = \frac{{}^{3}C_1}{{}^{51}C_1} = \frac{3}{51} = \frac{1}{17}$$

Since, cards are randomly selected without replacement

Hence, the probability of choosing two jacks will be:

$$P(A \cap B) = P(A) \times P(B)$$

$$P(A \cap B) = \frac{1}{13} \times \frac{1}{17}$$

$$P(A \cap B) = \frac{1}{221}$$

---

11. Basic rules of probability

Addition Rule 1: when two events, A and B, are mutually exclusive, the probability that A or B will occur is the sum of the probability of each event.

P(A or B) i.e. $P(A \cup B) = P(A) + P(B)$

Example

A day of the week is chosen at random. What is the probability of choosing a Saturday or Sunday?

Let the probability of choosing a Saturday be denoted as P(A).

$$P(A) = \frac{m}{n} = \frac{1}{7}$$

Let the probability of choosing a Saturday be denoted as P(B).

$$P(B) = \frac{m}{n} = \frac{1}{7}$$

Continued on the next page

The probability of choosing a Saturday or Sunday:

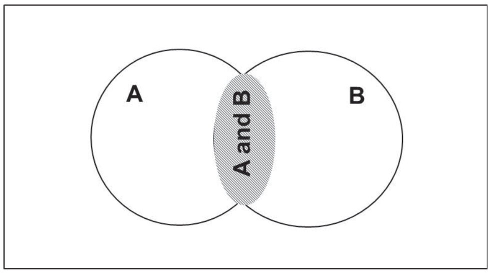
$P(A \cup B) = P(A) + P(B)$

$P(A \cup B) = \frac{1}{7} + \frac{1}{7}$

$P(A \cup B) = \frac{2}{7}$

Addition Rule 2: when two events, A and B, are not mutually exclusive, the probability that A or B will occur is:
P(A or B) i.e. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Non-mutually exclusive events can be presented as follows:



Tip

To find the probability of event A or B, we must first determine whether the events are mutually exclusive or non-mutually exclusive. Then we can apply the appropriate Addition Rule.

Example

In a class of 30 students, 13 are boys and 17 are girls. On a test, 4 girls and 5 boys achieved an A grade. If a student is chosen at random from the class, what is the probability of choosing a boy or choosing a student who scored an A?

Let the probability of choosing a boy be denoted as P (A).

$$P(A) = \frac{m}{n} = \frac{{}^{13}C_1 \times {}^{17}C_0}{{}^{30}C_1} = \frac{13}{30}$$

Let the probability of choosing a student who scored an A be denoted as P (B).

$$P(B) = \frac{m}{n} = \frac{{}^{4}C_1 + {}^{5}C_1}{{}^{30}C_1} = \frac{9}{30}$$

The probability of choosing a boy student who scored an A would be denoted as $P(A \cap B)$.

$$P(A \cap B) = \frac{m}{n} = \frac{{}^{5}C_1}{{}^{30}C_1} = \frac{5}{30}$$

The probability of choosing a boy or an A student would be denoted as $P(A \cup B)$

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

Since A and B are not mutually exclusive events

$P(A \cup B) = \frac{13}{30} + \frac{9}{30} - \frac{5}{30}$

$P(A \cup B) = \frac{17}{30}$

---

12. Independent events: two events, A and B, are independent if the fact that A occurs does not affect the probability of B occurring. An example of independent event is landing on heads after tossing a coin and rolling a 5 on a single 6-sided dice.

Multiplication rule 1: when two events, A and B, are independent, the probability of both occurring is:
P(A and B) = P(A) · P(B)

Example

A jar contains 3 red, 5 green, 2 blue and 6 yellow balls. A ball is chosen at random from the jar. After replacing it, a second ball is chosen. What is the probability of choosing a green and then a yellow ball?

Let the probability of choosing a green ball be denoted as P(A)

$$P(A) = \frac{m}{n} = \frac{{}^{5}C_1}{{}^{16}C_1} = \frac{5}{16}$$

Let the probability of choosing a yellow ball be denoted as P(B)

$$P(B) = \frac{m}{n} = \frac{{}^{6}C_1}{{}^{16}C_1} = \frac{6}{16}$$

Probability of choosing a green and then a yellow ball will be:

P (A ∩ B) = P (A) x P (B)

$$= \frac{5}{16} \times \frac{6}{16}$$

$$= \frac{15}{128} = 0.12$$

---

13. Conditional probability

Conditional probability of an event B is the probability that the event will occur given that event A has already occurred. In other words, event A has already happened, now what is the chance of event B happening? P(B|A) is also called the "Conditional Probability" of B, when A is given.

If P (A) > 0, then the probability of B given A is defined to be $P(B|A) = \frac{P(B \cap A)}{P(A)}$

Multiplication Rule 2: when two events, A and B, are dependent, the probability of both occurring is:
P (A ∩ B) = P (A) x P (B|A)

Example

A commerce teacher needs two volunteers to help him with a PowerPoint presentation for his class of 18 girls and 12 boys. He randomly chooses one student who comes to the front of the room. He then chooses a second student from those still seated. What is the probability that both students chosen are girls?

Let the probability of choosing a girl student be denoted as $P(A) = \frac{m}{n} = \frac{{}^{18}C_1 \times {}^{12}C_0}{{}^{30}C_1} = \frac{9}{15}$

The probability of B given A is defined to be $P(B|A) = \frac{m}{n} = \frac{{}^{17}C_1 \times {}^{12}C_0}{{}^{29}C_1} = \frac{17}{29}$

The probability that both students chosen are girls

P (A ∩ B) = P (A) x P (B|A)

$$P(A \cap B) = \frac{9}{15} \times \frac{17}{29}$$

$$P(A \cap B) = \frac{153}{435} = 0.35$$

Test Yourself 1

A number from 1 to 10 is chosen at random. What is the probability of choosing a 6 or an odd number?

A 1/2
B 1/5
C 3/5
D None of the above

Test Yourself 2

In a music school, 18% of all students play the piano and the guitar and 32% of all students play only the piano. What is the probability that a student plays the guitar given that the student plays the piano?

A 50.71%
B 58.12%
C 56.25%
D 14.75%

Let's look at a generic example of probability. This type of question may appear in your examination.

Example

A study circle of 8 members is to be formed from a group of 15 professionals comprising 9 men and 6 women.

What is the probability that the study circle would comprise:
(i) 3 women
(ii) At least 3 women

Answer

A circle of 8 members from 15 professionals can be formed as: ${}^{15}C_8$ OR $\frac{15!}{8! \times 7!}$

$$n= \frac{15 \times 14 \times 13 \times 12 \times 11 \times 10 \times 9 \times 8!}{8! \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}$$

n= 32432400/5040

n= 6435

(i) Let the probability of the study circle comprising 3 women be denoted as P(A). When a circle is formed with 3 women, the remaining 5 members need to be chosen from the 9 men. The 3 women can be selected in ${}^6C_3$ ways and the 5 men can be selected in ${}^9C_5$ ways.

Thus, event A comprising 3 women and 5 men can occur as: ${}^6C_3 \times {}^9C_5$

$$\text{Hence, } m = \frac{6 \times 5 \times 4}{1 \times 2 \times 3} \times \frac{9 \times 8 \times 7 \times 6 \times 5}{1 \times 2 \times 3 \times 4 \times 5} = 20 \times 126 = 2520$$

$$\text{Hence } P(A) = \frac{m}{n} = \frac{2520}{6435} = 0.39$$

Continued on the next page

(ii) Let the probability of the study circle comprising at least 3 women be denoted as P(B). At least 3 women means minimum 3 women should be in a circle; hence, we can have the following possible combinations:

Very important to interpret this!! Read tip below

(3 women and 5 men) OR (4 women and 4 men) OR (5 women and 3 men) OR (6 women and 2 men)

Hence,

$= ({}^6C_3 \times {}^9C_5) + ({}^6C_4 \times {}^9C_4) + ({}^6C_5 \times {}^9C_3) + ({}^6C_6 \times {}^9C_2)$

$= (20 \times 126) + (15 \times 126) + (6 \times 84) + (1 \times 36)$

$= 2520 + 1890 + 504 + 36$

$= 4950$

$$P(B) = \frac{m}{n} = \frac{4950}{6435} = 0.77$$

Tip

In the above example – part (ii), the requirement says 'at least' three women in a study circle. Accordingly, the possible combinations have been calculated. It is imperative for you as a student to know the use of 'and' and 'or' terms.

'And' denotes an intersection of events whereas 'or' denotes a union of events. In other words, when you interpret it as 'and', use the multiplication sign and when you interpret it as 'or', use the addition sign.

2. Calculate probability of events using binomial, Poisson and normal distribution.
   Apply probability concept in accounting and business situations.

[Learning Outcomes b and c]

Let us first understand the meaning of probability distribution. Probability distribution is a statistical function that describes all the possible values that a random variable can take within a given range and likelihood of a random variable occurring.. A probability distribution is a graph, a table, or a formula that gives the probability for each value of the random variable.

Random variable are of two types, discrete and continuous. It is a is a function that associates a unique numerical value with every outcome of an experiment. For example, while tossing a coin, the outcome of getting head or tail is regarded as random variable.

Expectation and variance of probability distribution

Expected value of probability distribution is calculated as: $E(x) = \frac{\sum X.p(x)}{n}$

Variance of probability distribution is calculated as $V(x) = \frac{\sum X^2.p(x)}{n} - \left(\frac{\sum X.p(x)}{n}\right)^2$

Example

On tossing a coin, there are two possible outcomes, head or tail. The probability distribution of number of head when one coin is tossed will be plotted as follows:

X	Probability of getting a head	Probability of getting a tail	Total probability
P (X)	0.50	0.50	1.00

We can see that for each value of x the probability is between 0 and 1. The sum of the probabilities is 1.

X (number of heads)	P(X)	X. P(x)	X^2. P (x)
0	0.50	0	0
1	0.50	0.50	0.50
	1.00	0.50	0.50

Continued on the next page

Expected value = E (x) = $\frac{\Sigma X.p(x)}{n}$ = 0.50/1 = 0.50

Variance = V (x) = $\frac{\Sigma X^2.p(x)}{n} - \left[\frac{\Sigma X.p(x)}{n}\right]^2$ = 0.50/1 – $(0.50/1)^2$ = 0.50 – 0.25 = 0.25

2.1 Binomial distribution

Bernoulli trial is a random experiment in which there are only two possible outcomes – success and failure.

1. Bernoulli Random Variables

A Bernoulli random variable X takes the values 0 and 1 and
P (X = 1) = p; P (X = 0) = 1 - p
It can be easily checked that the mean and variance of a Bernoulli random variable are E (X) = p and V (X) = p (1 - p)

Definition

Binomial Experiment: this is an experiment which satisfies the following four conditions.

1. There is a fixed number of trials i.e. 'n' number of trials.
2. Each trial is independent of the others.
3. There are only two outcomes; success and failure i.e. 'p' and 'q' (where, 'q' = 1 – 'p').
4. The probability of each outcome remains constant from trial to trial.

Hence, a binomial experiment is an experiment with a fixed number of independent trials, each of which can only have two possible outcomes. The fact that each trial is independent actually means that the probabilities remain constant.

Example

Examples of binomial experiments

a) Tossing a coin 20 times to determine the number of times that 'tails' will occur.
b) Asking 200 people if they watch TBC news.
c) Rolling a die to see if a 5 appears.

Examples which aren't binomial experiments

a) Rolling a die until a 6 appears (not a fixed number of trials)
b) Asking 20 people how old they are (not two outcomes)
c) Drawing 5 cards from a deck for a poker hand (done without replacement, so not independent)

---

2. Binomial probability function

The Binomial distribution is a discrete distribution expressing the probability of either success or failure.

The binomial distribution occurs in situations in which each of a number of independent trials (termed Bernoulli trials) results in one of two mutually exclusive outcomes. The mathematical description of the binomial distribution is as follows:

p (X) = ${}^{n}C_{x}\, p^{x}\, q^{n-x}$

Where,
p (X) = probability of success
n = number of Bernoulli trials
p = probability of success on any given trial
q = probability of failure on any given trial (i.e. 1 – p)
${}^{n}C_{x}$ = number of combinations of N things taken X at a time

Example

The probability of a cricketer hitting a four in an over is 0.3. If a cricketer remains not out during an over, what is the probability that he will hit a four 3 or fewer times? Note: assume that there is no extra bowl in that over.

Here,
X = 3 or fewer times, i.e. 0 or 1 or 2 or 3
n = 6 (an over has 6 bowls)
p = 0.3
q = 0.7

$p(X) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$
We need to find the probability that he will hit a four 3 or fewer times
= p (X = 0) + p (X = 1) + p (X = 2) + p (X = 3)
= 0.118 + 0.302 + 0.324 + 0.139
= 0.883

Where,
$p(X = 0) = {}^{6}C_{0}(0.3)^{0}\,(0.7)^{6} = 1 \times 1 \times 0.118 = 0.118$

$p(X = 1) = {}^{6}C_{1}(0.3)^{1}\,(0.7)^{5} = 6 \times 0.3 \times 0.168 = 0.302$

$p(X = 2) = {}^{6}C_{2}(0.3)^{2}\,(0.7)^{4} = 15 \times 0.09 \times 0.240 = 0.324$

$p(X = 3) = {}^{6}C_{3}(0.3)^{3}\,(0.7)^{3} = 20 \times 0.027 \times 0.343 = 0.185$

The following example will help you to determine the use of binomial probability distribution where you are asked to calculate probabilities involving analysis of linear equalities.

Example

If a student randomly guesses at five multiple-choice questions, find the probability that the student gets more than one correct. Each question has five possible choices.

Here, n = 5, p = 1/5, q = 4/5 and X > 1
The probability can be calculated as: P (X > 1) = P (X = 2) + P (X = 3) +P (X = 4) + P (X = 5)

Rather than calculating four probabilities, alternatively it can be calculated as: 1 – [P (X = 0) + P (X = 1)]

First, calculate the probability of getting one question or less wrong and then deduct that from 1. (Since in a probability distribution, total probability is always 1)

P (X > 1) = 1 – [P (X = 0) + P (X = 1)]
$P(X > 1) = 1 - [{}^{5}C_{0}(0.2)^{0}\,(0.8)^{5} + {}^{5}C_{1}(0.2)^{1}\,(0.8)^{4}]$
P (X > 1) = 1 – [(1 x 1 x 0.328) + (5 x 0.2 x 0.409)
P (X > 1) = 1 – [0.328 + 0.409]
P (X > 1) = 1 – 0.737
P (X > 1) = 0.263

Important

Analysis of linear qualities in probability

You need to memorise the following, as the exam questions may contain these symbols:
(a) $\geq$: greater or equal: At least, minimum of, no less than
(b) $\leq$: less or equal: At most, maximum of, no more than
(c) >: Is greater than, more than
(d) <: Is less than, smaller than, fewer than
(e) =: exactly

3. Mean and standard deviation

In a binomial distribution, it is simple to derive mean and standard deviation.

Formula of mean: $\mu = np$ (read as mew equal to n into p)

Formula of standard deviation (S.D.): $a = \sqrt{np\,(1-p)} = \sqrt{npq}$ (read as sigma equal to under root n into p into q)

Example

Let's take an example of tossing a coin. Determine the mean and the standard deviation for the number of heads obtained in 4 tosses.

Since there are only two outcomes (heads and tails) and the probabilities are constant (0.5 for both outcomes), the conditions for the binomial distribution are met.

Mean: $\mu = np = 4 \times 0.5 = 2$

Standard deviation: $a = \sqrt{npq} = \sqrt{4 \times 0.5 \times 0.5} = 1$

Test Yourself 3

The probability that goods will be delivered anywhere in Tanzania in two days or less is 0.84. If a company is sending consignments to 5 customers, and they are all due in 2 days, find the probability that 4 arrive within two days.

A 0.84
B 0.39
C 0.50
D 0.16

As the number of trials n in a binomial experiment increases, the probability distribution of the random variable X becomes bell-shaped. As a general rule of thumb, if $np\ (1 - p) > 10$, the probability distribution will be approximately bell-shaped.

Application of binomial distribution

Binomial distributions are regularly used in quality control systems. E.g. in the cases where an auditor tests random accounts out of a set of financial statements, and the audit assignment of financial statements is only accepted if the number of errors found among the sampled accounts falls below an identified number.

In order to analyse whether a sampling plan efficiently screens out financial statements containing a great number of errors and accepts financial statements with very few errors, one must calculate the probability, using binomial distribution, that financial statements will be accepted given various error levels.

Other common uses of binomial distributions in business include public opinion surveys, medical research and insurance problems. It can be applied to complex processes such as sampling items in factory production lines or estimation of percentage failure rates of products and components.

2.2 Poisson distribution

Binomial distribution can be used where the value of p (or q) is small and the sample size (n) is large. However, there are some practical difficulties with the calculations in the cases where n is too large. To overcome this situation, an alternative probability distribution model is used – the Poisson distribution.

The Poisson distribution was first derived in 1837 by a French mathematician Simeon Denis Poisson. The distribution arises when:

a) The events occur independently
b) The events occur such that the probability that two or more events occur simultaneously is zero.

Several books use different formulae for Poisson distribution; however, all give the same results. They are listed below:

$$P\ (X = x) = \frac{A^{x} e^{-A}}{x!}$$

Where,
e = 2.7183 (mathematical constant)
A = mean (parameter of the distribution)
x = 0, 1, 2...

Other formulae are:

$$P\ (X = x) = \frac{IJ^{x} e^{-IJ}}{x!}$$

Where,
e = 2.7183 (mathematical constant)
μ = mean (parameter of the distribution)
x = 0, 1, 2...

$$P\ (X = x) = \frac{m^{x} e^{-m}}{x!}$$

Where,
e = 2.7183 (mathematical constant)
m = mean (parameter of the distribution)
x = 0, 1, 2...

Important

The mean and variance of the Poisson distribution are equal.

Tip

Poisson distribution is a limiting form of binomial distribution in the following circumstances:

1. When n −7 CO
2. When p−7 0, q−71
3. When np = m

Example

Refer to the table that gives the number of days in a 50 day period during which electricity failure occurred in a factory. Fit a Poisson distribution to the data:

No. of electricity failures (in hour/s)	0	1	2	3	4
No. of days	19	18	8	4	1

Here, first we need to calculate the mean i.e. A.

No. of electricity failures (in hour/s)	No. of days	
X	f (X)	x. f (x)
0	19	0
1	18	18
2	8	16
3	4	12
4	1	4
	L:f (X) = n = 50	L:X.f (X) = 50

Continued on the next page

Mean: $A = \frac{L\ X.f(X)}{n} = \frac{50}{50} = 1$

Since the value of e^{-A} is not given, we need to derive it from the Poisson distribution table given in the appendices.

$e^{-1} = 0.3679$

Hence, $P\ (X = x) = \frac{A^x e^{-A}}{x!} = \frac{1^x\ (0.3679)}{x!}$

Example

A sole trader sorts the mail received each day into three types:

a) Financial (bills, communications from the bank, etc.),
b) Circulars (and all kinds of junk mail) and
c) Private (letters from friends, etc.).

The three types may each be modelled by independent Poisson distributions with means 1.1, 1.8 and 1.3 items of mail per day.

Find the probability that in any one day:

(i) the number of items of mail that are private or financial exceeds 3,
(ii) the total number of items of mail is less than 5.

(i) Private or financial mail

In this case the parameter is 1.1 + 1.3 = 2.4
From cumulative Poisson distribution tables,
P(number exceeds 3) = 1 – P(number is 3 or less)
$P\ (X > 3) = 1 - P(X \leq 3)$
= 1 – 0.7787
= 0.2213

(ii) All mail

In this case, the parameter is 1.1 + 1.8 + 1.3 = 4.2
From the tables, with A = 4.2, P(total number is less than 5)
= P(total number is less than or equal to 4) = 0.5898

Example

Motex is a textile mill which produces cotton materials. The number of defects in a cotton roll follows a Poisson distribution. The average number of defects in 50 metres of cloth is 1.2.

(i) Determine the probability of exactly three defects in 150 metres of cloth.
(ii) Determine the probability of at least two defects in 100 metres of cloth.
(iii) Determine the probability of exactly one defect in the first 50 metres of cloth and exactly one defect in the second 50 metres of cloth.

Here,

(i) Mean number of defects in 150 metres of cloth is 3.6. Hence, the probability of exactly three defects in 150 metres of cloth would be:

$P\ (X = 3) = \frac{A^x e^{-A}}{x!} = \frac{3.6^3\ e^{-3.6}}{3!} = \frac{46.656 \times 0.0273}{6} = 0.2123$

Continued on the next page

(ii) Mean number of defects in 100m of cloth is 2.4. Let X be the number of defects in 100 metres of cloth.

$$P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)]$$

$$= 1 - \left[\frac{2.4^0\, e^{-2.4}}{0!} + \frac{2.4^1\, e^{-2.4}}{1!}\right]$$

$$= 1 - \left[\frac{1 \times 0.0907}{1} + \frac{2.4 \times 0.0907}{1}\right]$$

$$= 1 - [0.0907 + 0.2177]$$

$$= 0.6916$$

(iii) Mean number of defects in 50m of cloth is 1.2. Let X be the number of defects in 50 metres of cloth.

$$P(X = 1) = \frac{A^x e^{-A}}{x!} = \frac{1.2^1\, e^{-1.2}}{1!} = \frac{1.2 \times 0.3012}{1} = 0.3614$$

Since X follows the Poisson distribution, the occurrence of defects in the first and the second 50 metres of cloth are independent. Hence, the probability of exactly one defect in the first 50 metres of cloth and exactly one defect in the second 50 metres of cloth = (0.3614) x (0.3614) = 0.1306

Application of Poisson distribution

The Poisson distribution is a discrete distribution. The major difference between Poisson and binomial distributions is that the Poisson distribution does not have a fixed number of trials. Instead, it uses a fixed interval of time or space in which the number of successes is recorded.

A typical, well-known, ancient and practical application of the Poisson distribution was approximating the annual number of Prussian horse regiment soldiers killed due to horse-kicks.

Poisson distribution is widely used as a statistical model for the number of events (such as the number of telephone calls at a business office, number of customers in waiting lines, number of defects in a given surface area, airplane arrivals, or the number of accidents at an intersection) taking place in a specific time period.

The application of the Poisson distribution enables managers to introduce optimal scheduling systems. For example, if the average number of people that eat in a certain restaurant on weekends is 500, then a Poisson distribution can answer such questions as, "What is the probability that more than 600 people will come to dine at that restaurant on a particular Saturday?"

It is also suitable in environmental studies, e.g., to model the number of prairie dogs found in a square mile of prairie. This distribution is widely used in physiology to determine the probabilistic occurrences of different types of neurotransmitter secretions.

Other recent examples include estimating the number of car crashes in a city of a given size; estimating the number of injuries in a typical cricket match, estimating the number of road accidents, etc.

Test Yourself 4

The mean number of defective products produced in a factory in one day is 3. What is the probability that in a given day there are exactly 5 defective products?

A 0.2
B 0.6
C Data inadequate
D None of the above

2.3 Normal distribution

If mean IJ = 0 and standard deviation a = 1, the distribution is called the standard normal distribution or the unit normal distribution, and a random variable with that distribution is a standard normal deviate. All normal distributions are symmetric and have bell-shaped density curves with a single peak.

A continuous random variable X is said to follow normal distribution with mean μ and standard deviation a, if its probability density function is:

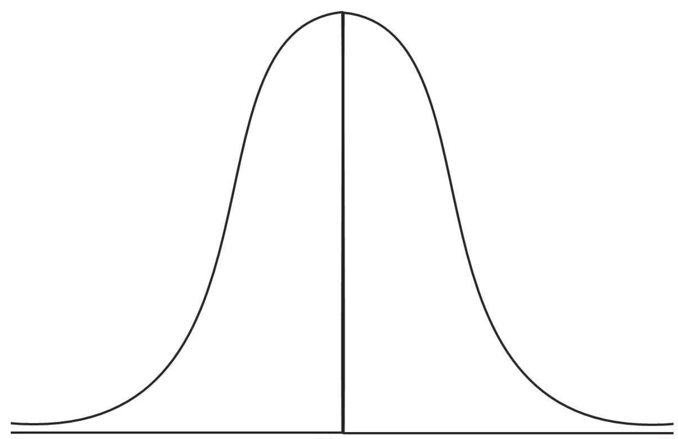
$$f(x) = \frac{1}{a\sqrt{2n}} e^{-\frac{1}{2}\left(\frac{x-IJ}{a}\right)^2}$$

The mean μ and standard deviation σ are called the parameters of normal distribution.

A normal distribution is a continuous probability distribution for a random variable x. The graph of a normal distribution is called the normal curve.

Normal probability curve

The curve representing the normal distribution is called the normal probability curve. It is always symmetric and bell shaped, as shown below:



A normal distribution has the following properties.

1. In this probability distribution, the value of mean, median and mode is equal i.e. $\overline{X}$ = M = Z
2. The normal curve is bell shaped. It is symmetric about the mean.
3. Under the normal curve, the total area is always equal to 1. The area under the curve to the right is equal to the area under the curve to the left i.e. 50%.
4. The normal curve approaches, but never touches, the x -axis as it extends farther and farther away from the mean.
5. According to the Empirical rule, approx. 68% of the area under the normal curve is between μ - a and μ + a; approx. 95% of the area under the normal curve is between μ - 2a and μ + 2a; and approx. 99.7% of the area under the normal curve is between μ - 3a and μ + 3a.
6. Amongst μ - a and μ + a, the curves of graph is drawn. The curves are drawn upward to the left side of μ - a and right side of μ + a. The points at which the curves change from upward to downward are known as inflection points.
7. The points of inflection are at x = μ ± σ
8. Since the curve is symmetrical, skewness = β1 = 0 and Kurtosis = β2 = 3.

Tip

Skewness refers to the symmetry of the curve.

a) If skewness = 0, distribution is symmetric and value of mean = value of median.
b) If skewness > 0 most values are concentrated on left of the mean, with extreme values to the right.
c) If skewness < 0 most values are concentrated on right of the mean, with extreme values to the left

Kurtosis refers to the degree of peak of the curve.

a) In case a normal distribution has kurtosis exactly 3 (excess kurtosis exactly 0). Any distribution with kurtosis =3 (excess = 0) is called mesokurtic.
b) In case, a distribution with kurtosis <3 (excess kurtosis <0) is called platykurtic. Compared to a normal distribution, its central peak is lower and broader, and its tails are shorter and thinner.
c) In case, a distribution with kurtosis >3 (excess kurtosis >0) is called leptokurtic. Compared to a normal distribution, its central peak is higher and sharper, and its tails are longer and fatter.

Tip

In standard normal distribution, mean (centre) is 0 and standard deviation (spread) is 1.

Standard normal distribution

Standard normal random variable, assuming that random variable X is normally distributed within mean μ and standard deviation a.

$$Z = \frac{X - \mu}{a}$$

Important

Area under curve

In case, you need to find the area under curve of Z = -1 to Z = +1; P ($\mu - \sigma < x < \mu + \sigma$) = 0.6826
In case, you need to find the area under curve of Z = -2 to Z = +2; P ($\mu - 2\sigma < x < \mu + 2\sigma$) = 0.9544
In case, you need to find the area under curve of Z = -3 to Z = +3; P ($\mu - 3\sigma < x < \mu + 3\sigma$) = 0.9973

This you can memorise as it may help you to solve your exam questions quickly.

Use of standard normal table to calculate probability

Tip

For calculating P (0 < Z < 1), in the table, probability 0.3413 should be referred to, which is in row 1.0 and column 0.00.

For calculating P (0 < Z < 1.25), in the table, probability 0.3944 should be referred to, which is in row 1.2 and column 0.05.

For calculating P (Z < 2.1), in the table, probability 0.4821 should be referred to, which is in row 2 and column 0.00 and should be added to 0.5000.

Similarly, for calculating P(Z > 3) = 0.5.0000 - 0.4986 = 0.0014.

Example

Let there be a normal random variable with mean 3 and variance 4. Compute the probability of P (-0.92 ≤ X ≤ 6.92)

Here,
μ = 3
a = 2 (square root of 4)

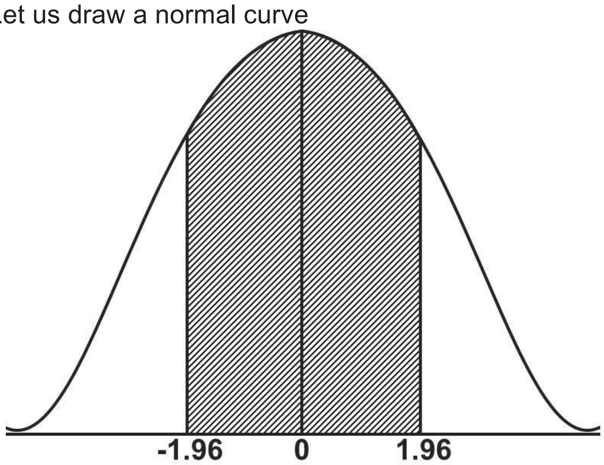
$$Z = \frac{X - \mu}{a}$$

$$Z = \frac{-0.92 - 3}{2} = -1.96$$

$$Z = \frac{X - \mu}{a}$$

$$Z = \frac{6.92 - 3}{2} = 1.96$$

Let us draw a normal curve



P (-0.92 ≤ X ≤ 6.92) = the value of Z score between -1.96 to 1.96
Referring to the Z score table:
P (-0.92 ≤ X ≤ 6.92) = 0.4750 + 0.4750 = 0.95

Let's study another example

Example

A survey was conducted to measure the height of Tanzanian females of the age group 21 - 30.The respondents' heights were normally distributed with a mean of 60 inches and a standard deviation of 5 inches. A study participant is randomly selected.

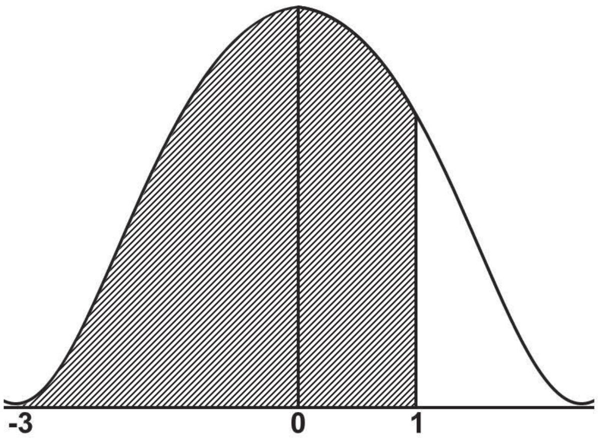
(a) Derive the probability of the female being between 45 and 65 inches.
(b) Derive the probability of the female being more than 60 inches tall.
(c) Derive the probability of the female being less than 56 inches tall.

Here, μ = 60 and a = 5

(a) The probability of the female being between 45 and 65 inches tall

$$Z = \frac{X - \mu}{a} = \frac{45 - 60}{5} = -3 \text{ and } Z = \frac{X - \mu}{a} = \frac{65 - 60}{5} = 1$$

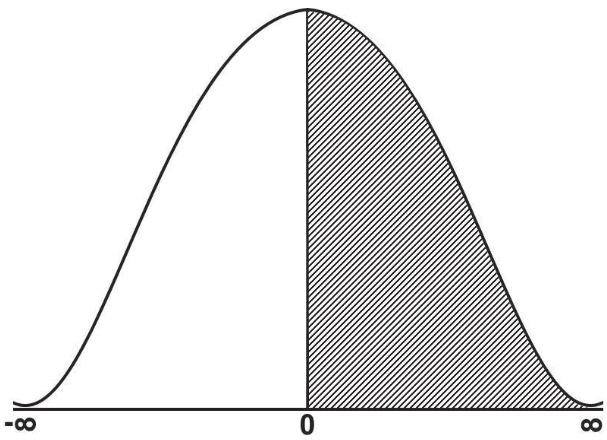
Continued on the next page



Hence, P (45 $\leq$ X $\leq$ 65) = P (-3 < Z < 1) = = P 3 < Z < 0) + P (0 < Z < 1) = 0.4986 + 0.3413 = 0.8399 i.e. 83.99% females are between 45 and 65 inches tall.

(b) The probability of the female being more than 60 inches tall

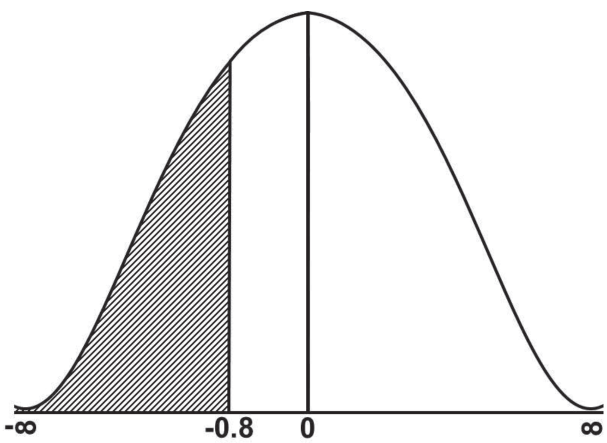
$$Z = \frac{X - IJ}{a} = \frac{60 - 60}{5} = 0$$



∴P(X > 60) = P (Z > 0) = P (0 < Z < ∞) = 0.5000 i.e. 50% females are more than 60 inches tall.

(c) The probability of the female being less than 56 inches tall

$$Z = \frac{X - IJ}{a} = \frac{56 - 60}{5} = -0.8$$



∴P(X < 56) = P (Z < -0.8) = P (0 < Z <- ∞) - P (0 < Z <-0.8) = 0.5000 – 0.2881 = 0.2119 i.e. 21.19% females are less than 56 inches tall.

Calculating mean and variance

Example

In a normal distribution, 31% of the items are under 45 and 8% are over 64. Find the mean and variance of the distribution.

Let x denotes the items that are given and it follows the normal distribution with mean μ and standard deviation a

The points X = 45 and X = 64 are located as shown in the figure.

(i) Since 31% of items are under X = 45, position of x into the left of the ordinate x = μ
(ii) Since 8% of items are above X = 64, position of this x is to the right of ordinate x = μ

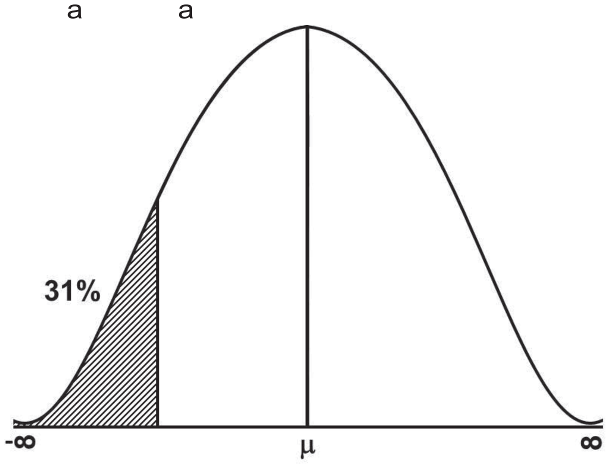
We need to derive two equations to calculate the value of mean μ and standard deviation a

For the first equation: when X = 45;

$$Z_1 = \frac{X - IJ}{a} = \frac{45 - IJ}{a}$$

For the second equation: when X = 64;

$$Z_2 = \frac{X - IJ}{a} = \frac{64 - IJ}{a}$$



As given in the question, P (X < 45) = 0.3100 and P (X > 64) = 0.08
Hence, P ($Z < -Z_1$) = 0.3100 and P ($Z > Z_2$) = 0.08

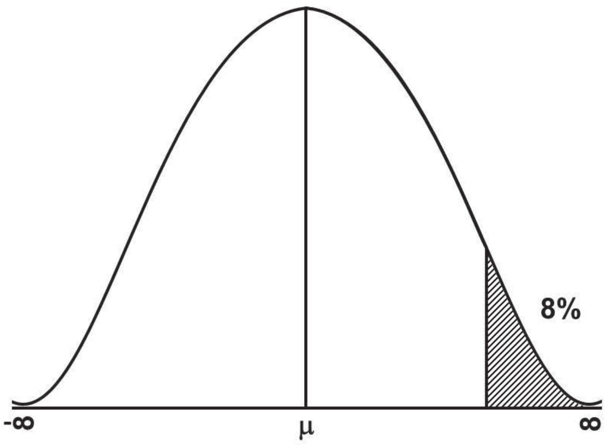
Based on the standard normal curve;

P ($Z < -Z_1$) = 0.3100
P ($- Z_1 < Z < 0$) = P ($- \infty < Z < 0$) – P ($-\infty < Z < Z_1$)
= 0.5000 – 0.3100
= 0.1900
Z_1 = 0.5 (from table) since the area is on the left side; Z_1 = -0.5

Similarly,

P ($Z > Z_2$) = 0.08
P ($0 < Z < Z_2$) = P ($0 < Z < \infty$) – P ($-Z_2 < Z < \infty$)
= 0.5000 – 0.0800
= 0.4200
Z_2 = 1.40 (from table)

Continued on the next page



Now elimination of equations is required to derive mean μ and standard deviation a

$Z_1 = -0.5 = \frac{45 - \mu}{a}$ and $Z_2 = 1.40 = \frac{64 - \mu}{a}$

: $-0.5a = 45 - \mu$ and $1.40a = 64 - \mu$
: $1.90a = 19$
: $a = 10$ and $\mu = 50$

Application of normal distribution

This is the most commonly perceived probability distribution. In the early 1800s, it was first used to analyse astronomical data by a German mathematician Karl Gauss.

In the fields of accounting and finance, normal distribution is used to aid forecast and adjust for a wide range of financial goals by optimising financial decision-making by applying and graphically mapping financial data into a distribution set of variables. It is used to make both quantitative and qualitative financial decisions which are based on mathematics.

It is also used in inventory management. The inventory forecasting model allows businesses to predict the highs and lows of their demands, which help them plan their operations efficiently for future inventory and customer demands.

It has many other applications too, such as for vehicle and health insurance analysis to assess the claims, reinsurance programs, etc. In the area of business administration, modern portfolio theory usually undertakes that the returns of a differentiated asset portfolio follow a normal distribution. Even in HRM, occasionally, the performance of the employees is considered to be normally distributed.

Test Yourself 5

In a certain departmental store, a supplier’s monthly statement shows an average balance of Tshs1200 million and a standard deviation of Tshs400 million. Assuming that the account balances are normally distributed:

(i) What percentage of the accounts is below Tshs1500 million?
(ii) What percentage of the accounts is over Tshs1500 million?
(iii) What percentage of the accounts is between Tshs1000 million and Tshs1500 million?

Answers to Test Yourself

Answer to TY 1

The correct option is C.

Let,
A = event of choosing number 6 (here, m = 1; since number 6 appears only once in the range from 1 to 10)
B = event of choosing an odd number (here, m = 5; since there are 5 odd numbers in the range from 1 to 10)

P (A) = $\frac{1}{10}$ and P (B) = $\frac{5}{10}$

The probability of choosing a 6 or an odd number (these are mutually exclusive events)

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

$P(A \cup B) = \frac{1}{10} + \frac{5}{10}$

$P(A \cup B) = \frac{6}{10} = \frac{3}{5}$

Answer to TY 2

The correct option is C.

Let
A = event that the student plays the piano
B = event that the student plays the guitar
Probability that the student plays the piano and the guitar: $P(A \cap B) = 0.18$
Probability that the student plays the piano: P (A) = 0.32
Probability that a student plays the guitar given that the student plays the piano: P(B|A) = ?

By applying the formula of conditional probability (i.e. multiplication rule 2):

$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.18}{0.32}$ = 0.5625 i.e. 56.25% students

Answer to TY 3

The correct option is B.

Here,

Event of success: consignment arrives within 2 days.
Event of failure: consignment does not arrive within 2 days.
Probability of success: p = 0.84
Probability of failure: q = 1 - p = 0.16
Number of trials (customers): n = 5
Number of successes: X = 4

$p(X) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$

$p(X = 4) = {}^{5}C_{4}(0.84)^{4}\,(0.16)^{1} = 5 \times 0.498 \times 0.16 = 0.398$

The probability that 4 of the 5 consignments will arrive on time is 0.398.

Answer to TY 4

The correct option is D.

Here,

Mean: A = 3
x = 5

Hence, P (X = 5) = $\frac{3^5 e^{-3}}{5!} = \frac{243\,(0.0497)}{120}$ = 0. 1008

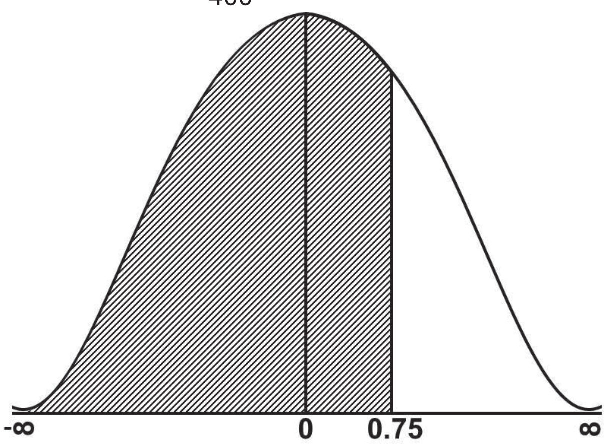
Answer to TY 5

The correct option is B.

Here, mean μ = 1200 and standard deviation a = 400

The percentage of accounts below Tshs1500 million

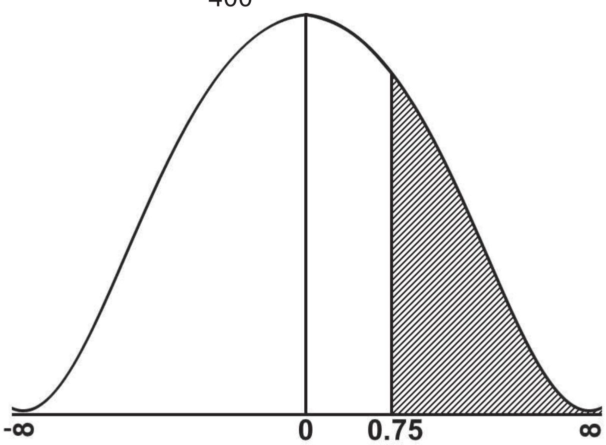
$Z = \frac{X - IJ}{a} = \frac{1500 - 1200}{400} = 0.75$



P (X < 1500) = P (Z < 0.75)
P (-co < Z < 0.75) =P (-∞ < Z < 0) + P (0 < Z < 0.75)
P (-co < Z < 0.75) = 0.5000 + 0.2734= 0.7734 i.e. 77.34%

The percentage of accounts over Tshs1500 million

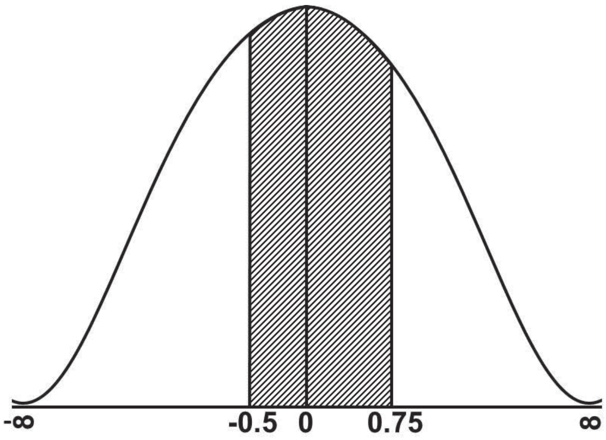
$Z = \frac{X - IJ}{a} = \frac{1500 - 1200}{400} = 0.75$



P (X > 1500) = P (Z > 0.75)
=P (0 < Z < ∞) – P (0.75 < Z < ∞)
= 0.5000 – 0.2734= 0.2266 i.e. 22.66%

The percentage of accounts between Tshs1000 million and Tshs1500 million

$$Z = \frac{X - \mu}{\sigma} = \frac{1000 - 1200}{400} = -0.5$$



P (1000 < X < 1500) = P (-0.5 < Z < 0.75) = 0.1915 + 0.2734 = 0.4649 i.e. 46.49%

Self Examination Questions

Question 1

A shelf in a garments shop contains 10 pink scarves, 10 yellow scarves, 10 olive scarves and 10 navy scarves. You, as a customer, empty the shelf on the table.

What is the probability of grabbing out two scarves of the same colour in a row for you and your friend?

A 1/40
B 9/39
C 2/9
D 8/37

Question 2

In a review meeting, team X hides facts 30% of the time and team B reveals facts 80% of the time. Determine the probability that both team X and team Y are contradicting each other.
-
A 94%
B 6%
C 56%
D 38%

Question 3

A company launched a new website. According to its analysis, the number of visitors to a webserver per minute follows the Poisson distribution.

Assuming the average visitors per minute as 4, determine the probability of two or more persons visiting in a minute.

A 0.908
B 0.238
C 0.018
D 0.574

Question 4

Sahara Engineering acquires new machinery. The probability that the new machinery will produce a defective item is 20%.

If a random sample of 6 items is taken from the output of the new machinery, what is the probability that there will be 5 or more defective items in the sample?

A 20%
B 80%
C 16%
D None of the above

Question 5

Derive the mean and standard deviation of marks in an interview where 44% of the candidates obtained marks below 55 and 6% scored marks above 80.

Answers to Self Examination Questions

Answer to SEQ 1

The correct option is B.

Let the probability of choosing two scarves of the same colour be denoted as P (A).

Choosing either two pink scarves OR two yellow scarves OR two olive scarves OR two navy scarves from total 40 scarves

$$P(A) = \frac{m}{n}$$

$$P(A) = \frac{{}^{10}C_2}{{}^{40}C_2} + \frac{{}^{10}C_2}{{}^{40}C_2} + \frac{{}^{10}C_2}{{}^{40}C_2} + \frac{{}^{10}C_2}{{}^{40}C_2}$$

$$P(A) = \frac{9}{156} + \frac{9}{156} + \frac{9}{156} + \frac{9}{156}$$

$$P(A) = \frac{36}{156} = \frac{9}{39}$$

Answer to SEQ 2

The correct option is D.

Let
A event = Team X hides facts; P (A) = 0.30
A' event = Team X reveals facts; P (A') = 0.70
B event = Team Y hides facts; P (B) = 0.20
B' event = Team Y reveals facts; P (B') = 0.80

Contradictions takes place when one team hides facts and another team reveals facts.
i.e. $P(A \cap B') + P(A' \cap B)$
= (0.30 x 0.80) + (0.70 x 0.20)
= 0.24 + 0.14
= 0.38
= 38%

Answer to SEQ 3

The correct option is A.

Probability of two or more persons visiting = P $(X \geq 2)$
Mean: A = 4

$P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)]$

$$P(X \geq 2) = 1 - \left[\frac{4^0 e^{-4}}{0!} + \frac{4^1 e^{-4}}{1!}\right]$$

$$P(X \geq 2) = 1 - \left[\frac{1 \times 0.0183}{1} + \frac{4 \times 0.0183}{1}\right]$$

$P(X \geq 2) = 1 - [0.0183 + 0.0.732]$
$P(X \geq 2) = 0.9085$

Answer to SEQ 4

The correct option is C.

$p(X) = {}^nC_x\ p^x\ q^{n-x}$

here,
X = 5 or more i.e. 5, 6
n = 6
p = 0.20
q = 0.80

$p(X \geq 5) = p(X = 5) + p(X = 6)$

$p(X \geq 5) = {}^6C_5\ (0.20)^5\ (0.80)^1 + {}^6C_6\ (0.20)^6\ (0.80)^0$
= (6 x 0.00032 x 0.8) + (1 x 0.000064 x 1)
= 0.0015 + 0.000064
= 0.001564
= 0.0016
= 16%

Answer to SEQ 5

We need to derive two equations to calculate the value of mean μ and standard deviation a: 44% of the candidates obtained marks below 55 and 6% scored above 80.

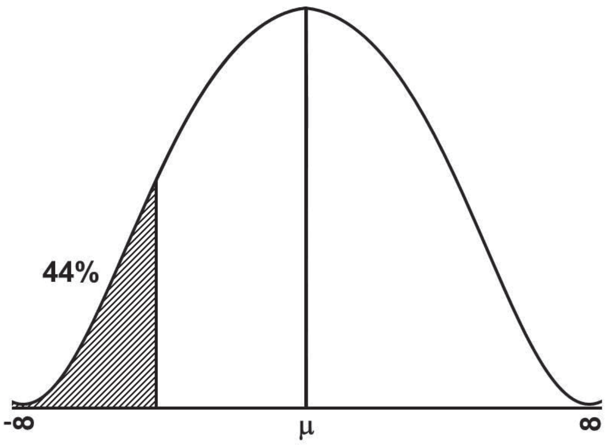
For the first equation: when X = 55;

$$Z_1 = \frac{X - \mu}{a} = \frac{55 - \mu}{a}$$

For the second equation: when X = 80;

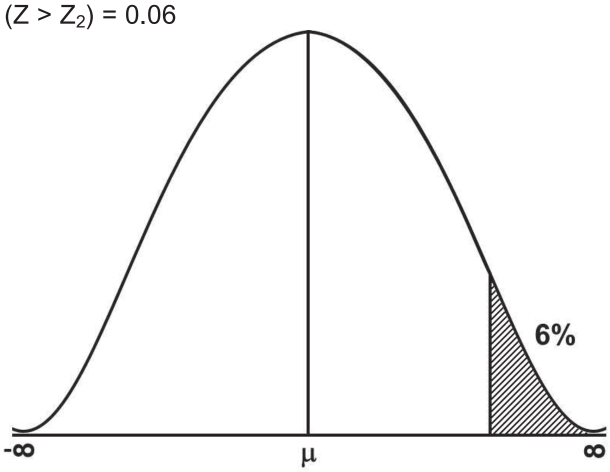
$$Z_2 = \frac{X - \mu}{a} = \frac{80 - \mu}{a}$$

As given in the question, P (X < 55) = 0.4400 and P (X > 80) = 0.06
Hence, P $(Z < -Z_1)$ = 0.4400 and P $(Z > Z_2)$ = 0.06



Based on the standard normal curve;
$P(Z < -Z_1) = 0.4400$
$P(-Z_1 < Z < 0) = P(-\infty < Z < 0) - P(-\infty < Z < Z_1)$
$= 0.5000 - 0.4400$
$= 0.0600$
$Z_1 = 0.15$ (from table) since the area is on the left side $Z_1 = -0.15$

Similarly,
$P(Z > Z_2) = 0.06$



$P(0 < Z < Z_2) = P(0 < Z < \infty) - P(-Z_2 < Z < \infty)$
$= 0.5000 - 0.0600$
$= 0.4400$
$Z_2 = 1.55$ (from table)

Now elimination of equations is required to derive mean μ and standard deviation a

$Z_1 = -0.15 = \frac{55 - IJ}{a}$ and $Z_2 = 1.55 = \frac{80 - IJ}{a}$

: $-0.15a = 55 - \mu$ and $1.55a = 80 - \mu$
: $1.70\sigma = 25$
: $a = 14.71$ and $\mu = 57.21$

5

RANDOM VARIABLE, PROBABILITY DISTRIBUTION AND ELEMENTS OF DECISION ANALYSIS

Get Through Intro

Decision making requires us to choose one course of action from many. This is where the difficulty starts. How can you be sure that your choice is the right one? Part of the accountant’s role is to try to reduce the risk associated with decision making by evaluating the options.

The use of various mathematical formulae can help to choose the best alternative. One of the methods that can be used is the “expected value” method. The possible outcomes of any action taken are quantified under this approach to arrive at an “expected value” which presents us with a probable outcome of a decision.

This Study Guide will take us through the detailed methods of the calculations of the “expected value” under different situations.

A management accountant has to take business decisions on, for example, a change in the selling price of a product, altering the product mix or the launch of a new product, which are critical and expose the organisation to risk. There are a multitude of choices available in the course of decision making. The “expected value” approach is one way to evaluate such problems. This Study Guide prepares you for having to use these techniques in real life.

Learning Outcomes

a) Determine the expected value of a random variable.
b) Make decision under uncertainty and under risk.
c) Determine expected values of perfect information.
d) Construct a decision tree.
e) Apply a decision tree in decision making.
f) Define central limit theorem and use it in the concept of sample means.
g) Explain the appropriate sampling distributions of the sample means and sample proportions.
h) Apply the concept of decision analysis in accounting and business situation.

1. Determine the expected value of a random variable.
[Learning Outcome a]

The expected value (or expectation) of a random variable is the sum of the payoff ('value') of each possible outcome of an experiment multiplied by the respective probability.

The calculation of the expected values is essential as the results give us the consequences that could emerge from a decision taken .It represents the average amount one expects to win per bet if the bet is repeated many times with identical odds. It should be noted that the value itself might not be expected in the general sense; it may be unlikely or even impossible. Thus the actual win or loss may never be equal to expected value. However if the bet is repeated several times then on average this is what you would win each time.

The very substance of finding the expected values emanates from the fact that we have to take decisions that are subject to various outcomes. Decision making is one of the prime responsibilities of management at any level. Decision making is the managerial process of selecting the alternative that is most likely to attain the desired outcome from available choices.

It involves committing the organisation and its resources to specific courses of action. All managers have to take decisions and solve problems.

Uncertainty and problems arise when the actual state of affairs differs from the desired. Problems may also give rise to favourable opportunities. Decision making involves bridging the gap between the situation as it stands and the desired situation through problem solving and creating opportunities.

Now let us try to calculate "expected value" with the help of the following simple example.

Example

Susan, freelance interior designer, has been saving her pocket money. She has managed to save Tshs200 million and would now like to put it to good use. It has been suggested to her that she invest the money in the shares of Union Plc, but she is unsure whether it will give her a good return. The following information is available regarding the performance of the shares:

Rate of return	Probability
10%	0.20
12%	0.30
9%	0.50

Susan would like to calculate the return that she can expect from investing in these shares.

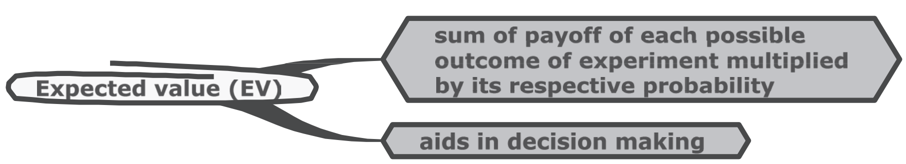
This can be calculated as follows
For all possible outcomes,

Sum of [Probability of getting a particular value x value (also known as 'pay-off')]

In this case it will be:

For all possible rates of return on investment
Sum of [(Probability of getting a particular rate of return) x (expected value of the return, that is the interest earned)]
= ((Tshs200 million x 10%) x 0.20) + ((Tshs200 million x 12%) x 0.30) + ((Tshs200 million x 9%) x 0.50))
= Tshs20.20
i.e. Susan can expect a return of Tshs20.20 million by investing Tshs200million in Union Plc.

SUMMARY



1.1 Expected values

Expected value is the weighted average of the alternative payoffs estimated for different states of nature i.e. the events. The weighted average of the payoffs is made with reference to the related probabilities for the possible states of nature.

Expected value is the sum of the probability of each possible outcome of the strategy multiplied by its payoff.

Example

Harry wants to catch a flight. The strategies before him are either to go to the airport by bus or to go by cab. Either he catches the flight or he misses it. These are the events.

If he goes by bus, the probability of catching the flight is 0.90 but if he goes by cab, the probability of catching the flight is 0.99. The bus will cost him Tshs2,000 and a cab, Tshs5,000. If he catches the flight, he will finalise a business deal that will fetch him a profit of Tshs150,000.

He will calculate the expected cost of going by bus as well as the cost of going by cab and will choose the option that will give him the greater gain in the given situation.

While calculating expected value, it is necessary to consider all the possible events related to the strategies.

1. Important terms

Some important terms related to the expected value:

(a) Events

Events are the occurrences beyond the control of the decision-maker. The events are often called outcomes or states of nature.

Example

Continuing the above example

Catching the flight or missing it, are the events.

(b) Payoff

Payoff is a conditional value of a strategy employed by a decision-maker. It indicates a net gain or net loss for the decision-maker as a result of the strategy adopted by him against a state of nature.

When the strategies and the events are arranged in a tabular format, the outcome of each strategy as against each event is written in the cell where the rows (usually signifying the events) intersect with the columns (usually signifying the strategies).

Example

Continuing the above example

Payoff can be calculated as follows: if he catches the flight, the gain will be (150 - 2) for the bus and (150 - 5) for the cab, if he misses the flight, the loss will be 2 for the bus and 5 for the cab.

State of nature	Strategies	
	Going by Bus (Tshs'000)	Going by Cab (Tshs'000)
Catches the flight	148	145
Misses the flight	(2)	(5)

Every action has as many payoffs as the events i.e. if there are 3 events related to a strategy; the number of payoffs will be 1 x 3 = 3. If the strategies are 3 and the numbers of possible events are 3, the number of payoffs will be 3 x 3 = 9.

(c) Strategy

Strategy is a course of action to be chosen by a decision-maker.

Example

Continuing the above example

Strategies before Harry are either to go to the airport by bus or to go by cab.

---

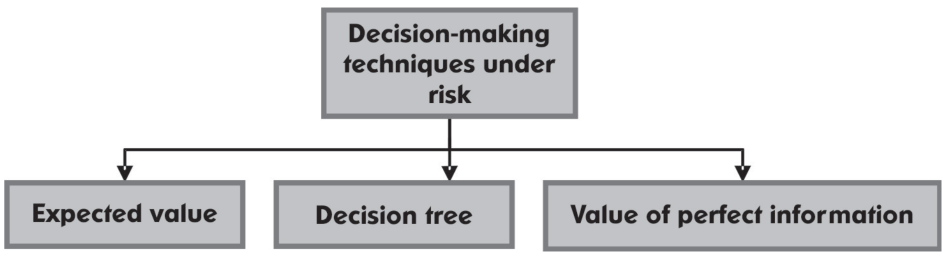
2. Application of expected value

Expected value is widely-used in situations involving risk, in order to evaluate the alternatives. While taking a decision under risk, the decision-maker, on the basis of knowledge, experience or judgement, assigns probabilities to the likelihood of an occurrence of a state of nature.

Steps to apply expected value technique in solving decision-making problem:

1. Identify strategies and states of nature
2. Calculate payoffs
3. Calculate EMV (economic monetary value) for each strategy
4. Choose the strategy with the maximum expected value

Diagram 1: Decision-making techniques under risk



In the expected value technique, different alternatives are evaluated in order to choose the one that presents the maximum gain. Expected value is normally expressed in a tabular format, where the events (states of nature) are arranged in the far left column, and the acts (strategies) in the topmost row.

The form of a payoff table is produced below:

State of nature	Probabilities	Strategies			
		S_1	S_2	S_3	S_4
N_1		x_{11}	x_{12}	x_{13}	x_{14}
N_2		x_{21}	x_{22}	x_{23}	x_{24}
N_3		x_{31}	x_{32} Payoffs	x_{33}	x_{34}
N_4		x_{41}	x_{42}	x_{43}	x_{44}

3. Calculation of expected value

$$\text{Expected value } [E(s)] = p_1x_{11}+p_2x_{12}+\ldots\ldots+p_{n-1}x_{1n-1}+p_nx_{1n}$$
$$= \sum_{j=1}^{n} p_j x_{ij}$$

Where,
's' denotes a particular strategy
'x_{ij}' denotes payoffs at the juncture of strategy j and state of nature i
'n' denotes number of possible events for a particular strategy
The probabilities of the possible events are denoted by p_1, p_2,p_{n-1},p_n.

4. Application of the expected values technique in solving decision-making problems

Example

Praise, a car dealer, provides the following information about a car (model Comfort) it trades -

Daily demand for Comfort	1	2	3	4
Probability of demand	0.2	0.3	0.4	0.1

The cost of Comfort is Tshs12,500,000 and its selling price is Tshs15,000,000.

Calculating the expected value, advice the company on how many cars it should purchase daily.

Answer

Step 1

Identify strategies and states of nature

Here, the strategies would be:
S_1 Buy one car
S_2 Buy two cars
S_3 Buy three cars
S_4 Buy four cars

The states of nature are the demand for the car.

Step 2

Calculate payoffs

Accordingly we will calculate the payoff for each strategy corresponding to the states of nature. Payoff will be calculated as:

Payoff = Selling price of the car – (Cost of the car x No. of cars)

Payoff table in (Tshs million)

States of nature	Probabilities	Strategies			
		S_1	S_2	S_3	S_4
N_1	0.2	2.5	-10(W1)	-22.5	-35
N_2	0.3	2.5	5	-7.5	-20
N_3	0.4	2.5	5	7.5	-5
N_4	0.1	2.5	5	7.5	10

Step 3

Calculate EMV for each strategy

We will calculate the EMV for each strategy

$S_1 = 0.2(2.5) + 0.3(2.5) + 0.4(2.5) + 0.1(2.5) = 2.5$

$S_2 = 0.2(-10) + 0.3(5) + 0.4(5) + 0.1(5) = 2$

$S_3 = 0.2(-22.5) + 0.3(-7.5) + 0.4(7.5) + 0.1(7.5) = -3$

$S_4 = 0.2(-35) + 0.3(-20) + 0.4(-5) + 0.1(10) = -14$

Continued on the next page

Step 4

Choose the strategy with the maximum expected value

Here, the maximum expected value is for strategy S_1. Therefore, Praise should be advised to buy 1 car a day.

Workings

W1 Calculation of payoff

If strategy 2 i.e. buy two cars is adopted in states of nature N1 where the demand is for only one car, the payoff will be:

Payoff = (Tshs15,000,000 x 1 car) – (Tshs12,500,000 x 2 cars)
= Tshs(10,000,000)

As the cost incurred for buying cars is more than the sales revenue, a loss of Tshs10,000,000 will be incurred.

The payoff for other strategies corresponding to each state of nature can be calculated in the same manner.

Test Yourself 1

A company is considering whether to take up two new projects with the following probabilities of making profits.

Project	Profit (Tshs'000)	Probability
Project A	2,000	0.6
	3,000	?
Project B	5,000	0.4
	4,000	0.3
	6,000	?

What is the probability of project A and B making a profit of Tshs3,000,000 and Tshs6,000,000 respectively?

A 0.4, 0.6
B 0.5, 0.3
C 0.4, 0.3
D 0.5, 0.5

2. Make decision under uncertainty and under risk.
   Determine expected values of perfect information.

[Learning Outcomes b and c]

Expected values aid decision making, but unless we know how to use them effectively as tools in this process, they are of no use. When there is only one alternative the outcome is certain and predictable. This is decision making under certainty.

However in most cases the decisions are subject to uncertainty. The major difficulty in such cases is the choice of the best from amongst many. The outcomes of alternatives are subject to many possible events (or states of nature). Now since we have to make good decisions we follow the pay off table approach to decision making.

A payoff table is a tabular representation of the outcomes of the alternatives under the given possible events. To construct a payoff table:

1. Arrange the events (states of nature) in the leftmost column.
2. Arrange the acts (alternatives) in the topmost row. This will divide the table as one row for each event and one column for each act.
3. Once this matrix structure is made, fill in the outcomes of each act as against each event in the cell where they intersect.

To gain a better understanding let us try to construct a payoff table for the example below.

Example

BSR plc manufactures bikes. It wants to launch a new model in the quarter before Christmas as the demand is high during this time. The company has three models lined up for launch; Legend, Invader and Street Hawk. BSR plc naturally wants to launch the most profitable bike in the season. The demand conditions as forecasted by the company are:

Demand condition	Legend	Invader	Street hawk
High	200 units	150 units	80 units
Moderate	150 units	90 units	60 units
Low	100 units	70 units	10 units

The profits associated with each bike are Legend Tshs20,000, Invader Tshs35,000 and Street Hawk Tshs50,000.

The three demand conditions given above form the 'events'. The acts in this case constitute the launch of any of the three bikes, Legend, Invader or Street Hawk. As the demand conditions and profits are different for all bikes the outcomes will differ for each condition. The payoff table will be as follows:

Acts → / Events ↓	Launch Legend (Profit 20) Tshs'000	Launch Invader (Profit 35) Tshs'000	Launch Street Hawk (Profit 50) Tshs'000
High demand	200 x 20 = 4,000	150 x 35 = 5,250	80 x 50 = 4,000
Moderate demand	150 x 20 = 3,000	90 x 35 =3,150	60 x 50 = 3,000
Low demand	100 x 20 = 2,000	70 x 35 = 2,450	10 x 50 = 500

This forms our payoff table arranged in a matrix structure. By assigning probabilities to the demands we can formulate an expected value table and ascertain the best option. The probabilities of the demand forecast of high, moderate and low are 0.1, 0.7 and 0.2 respectively as provided by the finance manager, based on market research.

Continuing the same format as the table above, we multiply the probability with each payoff and sum up the products for each act. The expected value is normally measured in monetary units and hence also called 'Expected Monetary Value' (EMV). When we multiply the respective pay offs with their probabilities, we attain an expected value for each pay off.

COMPUTATION OF EMV OF VARIOUS ACTS

Acts → / Events ↓	Probability	Launch Legend Tshs'000	Launch Invader Tshs'000	Launch Street Hawk Tshs'000
High demand	0.1	4,000 x 0.1 = 400	525	400
Moderate demand	0.7	3,000 x 0.7 = 2,100	2,205	2,100
Low demand	0.2	2,000 x 0.2 = 400	490	100
EMV		2,900	3,220	2,600

The product of the payoffs and probabilities gives us conditional values hence the table is also called a 'conditional value table'. The sum of the pay offs for each act gives us the EMV of the chosen act. Since we want to maximise profit we choose to launch Invader as it has the maximum EMV.

Alternatively, the result can also be arrived at by using the concept of opportunity loss. The approach is called Expected Opportunity Loss (EOL). Opportunity loss is the difference between the highest possible profit for an event and the actual profit for an action taken. To get this we simply subtract each payoff in each row from the highest payoff for that row (since the events are arranged in a row). The opportunity loss table or regret table thus obtained gives us the values indicating opportunity loss due to failure to take the best action.

Example

Continuing with the previous example

Converting the above given payoff table to a regret table we get:

Acts → / Events ↓	Launch Legend (Profit 20) Tshs'000	Launch Invader (Profit 35) Tshs'000	Launch Street Hawk (Profit 50) Tshs'000
High demand	5,250 - 4,000= 1,250	5,250 - 5,250 = 0	5,250 - 4,000 = 1,250
Moderate demand	3,150 - 3,000 =150	3,150 - 3,150 = 0	3,150 - 3,000 = 150
Low demand	2,450 - 2,000 = 450	2,450 - 2,450 = 0	2,450 - 500 = 1,950

To ascertain the EOL we assign probabilities to each regret or loss value obtained:

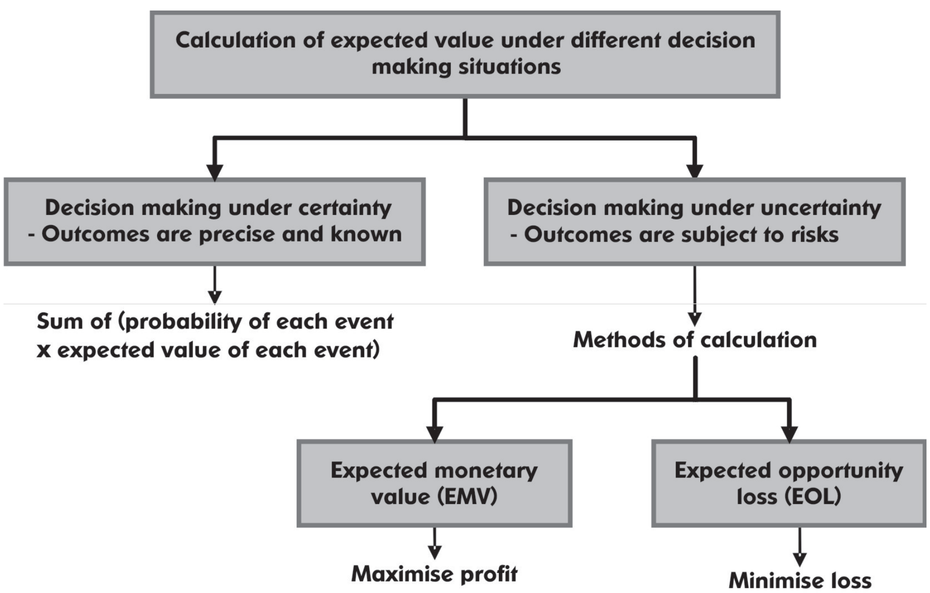
Acts → / Events ↓	Probability	Launch Legend Tshs'000	Launch Invader Tshs'000	Launch Street Hawk Tshs'000
High demand	0.1	1,250 x 0.1 = 125	0 x 0.1 = 0	1,250 x 0.1 = 125
Moderate demand	0.7	150 x 0.7 = 105	0 x 0.7 = 0	150 x 0.7 = 105
Low demand	0.2	450 x 0.2 = 90	0 x 0.2 = 0	1,950 x 0.2 = 390
EOL		320	0	620

The aim here is to minimise the regret or the EOL. Hence we choose the act with the minimum EOL which is to launch Invader.

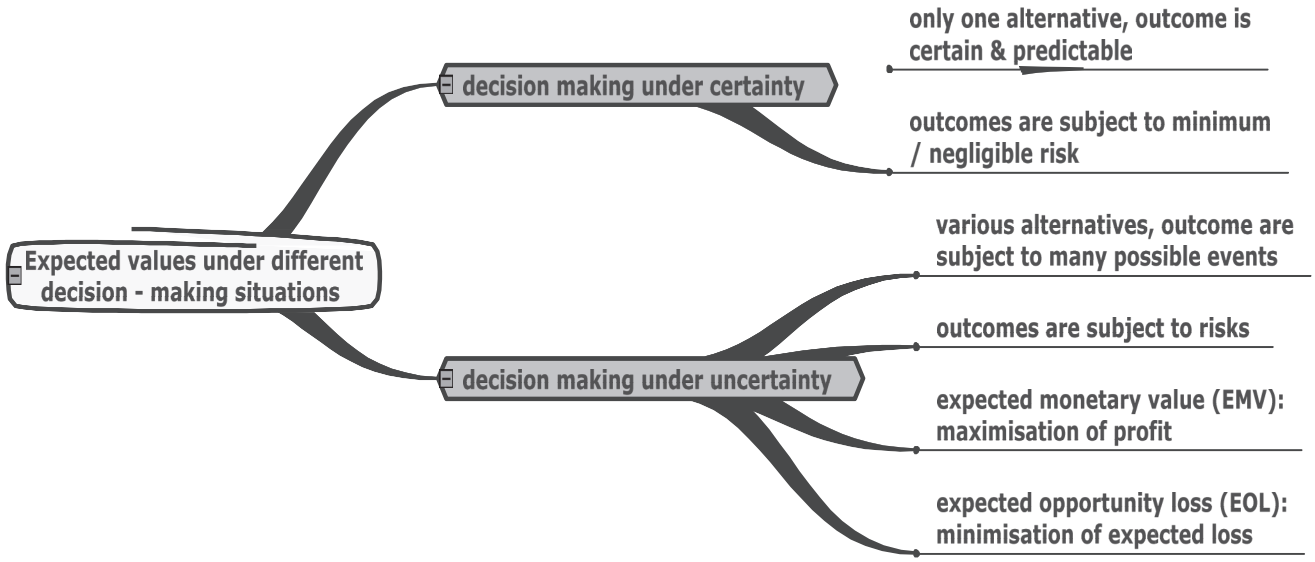
Though the EMV and the EOL criteria are conceptually different, in practice both the criteria will lead to choosing the same strategy. Hence the strategy with the maximum EMV is also the strategy with the minimum EOL.

Expected values can be used in various simple decision making problems in the given manner using either EMV or EOL approach.

Diagram 2: Methods of calculation of expected value



SUMMARY



Test Yourself 2

The payoff table is also known as:

A Conditional value table
B Conditional table
C Values table
D Optional value table

Test Yourself 3

A pay-off table is given below

Events	Acts		
	A_1 Tshs'000	A_2 Tshs'000	A_3 Tshs'000
E_1	25	50	80
E_2	30	60	90
E_3	40	70	100

What will be the opportunity loss values for the above payoffs corresponding to event E_1?

A Tshs55,000, Tshs30,000, Tshs0
B Tshs60,000, Tshs30,000, Tshs0
C Tshs60,000, Tshs30,000, Tshs0
D None of the above

Test Yourself 4

The alternatives under decision making are called:

A Acts
B Events
C Pay offs
D Expected values

Test Yourself 5

Helliz Plc has found that there is a 45% chance of making a profit of Tshs120 million and a 55% chance of incurring a loss of Tshs80 million the next year. What is the expected value of profit / loss for the next year?

A Tshs30 million loss
B Tshs30 million profit
C Tshs10 million loss
D Tshs10 million profit

Test Yourself 6

Which of the following is correct for expected monetary value?

A It is computed using a decision table with conditional values for all events or states of nature
B It will be determined for each alternative as the sum of possible payoffs of the alternative, each weighted by the probability of that payoff occurring
C It measures conditional values (payoffs) in monetary units
D All of the above

Every mathematical technique suffers from some limitations and the “expected value technique” is no exception. Some possible limitations of this technique are explained below.

1. Decision making under risk

Definition

Risk

a) A situation that could be dangerous or have a bad outcome.
b) The possibility that something unpleasant will happen.
c) An exposure to danger or loss.

Due to the risk factor, the decisions based on this risk are very hard to rely on. The positive outcome and the negative outcome are equally likely when they are surrounded by risk. In such a case where the outcomes are so unlikely, placing a definite value on the outcome is totally meaningless. The expected values are hence unreliable.

Example

The decision of number of units to be produced is taken based on the expected demand. This decision can be taken if we compute the expected demand based on the probability. However the probabilities are subject to risk. If the expected demand comes to 1,500 units and we produce these, but the actual demand turns out to be only 750 units then the remaining units will be excess production that can become obsolete.

---

2. Assigning probability values

Definition

A probability is a number in between 0 and 1, both inclusive, indicating the likelihood of the occurrence of an outcome.

Expected values based on these probabilities also become subject to the judgement of the decision maker. In addition to judgement, reliability of information, knowledge and experience of the decision maker also affect the authenticity of the expected values.

Example

In the above example the probability based on which the expected demand was computed, provided us with a wrong answer. The incorrect probabilities in this case resulted in excess production of 750 units. Accuracy of the probabilities is hence an essential factor that decides the precision of expected values.

3. Choice of alternative

The alternative chosen should be the one with the highest EMV (expected monetary value). However, this alternative could be laden with immense risk as well. So it is in the best interests of all concerned to select this alternative?

Expressed differently, the best alternative could also be the one with maximum risk. Hence, EMV is not a reliable basis for decision making under uncertainty.

Example

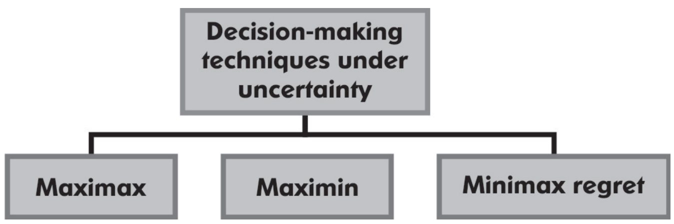
In the above example the chosen alternative could be the one with the highest EMV. However this decision also failed as the demand turned out to be just 50% of the expected demand. The actual demand is highly subjective to the prevalent conditions and as such choice of the alternative with the highest EMV is risky.

4. The time value of money is not taken into consideration.

2.1 Techniques of maximax, maximin and minimax regret to decision-making problems including the production of profit tables

Maximax, maximin and minimax regret are decision-making techniques adopted in situations of uncertainty. They are suitable techniques to be used under uncertainty due to unavailability of historical data, lack of experience, etc.

Diagram 3: Decision-making techniques under uncertainty



1. Maximax

This is an optimistic approach to decision-making. Leonid Hurwitz suggested this decision criterion. Under this approach, the alternative that maximises the maximum outcome for every alternative strategy is selected.

The steps followed in this approach:

a) Find out the maximum payoff for each of the strategies.
b) Select the maximum out of the maximum payoffs for each strategy as identified in the preceding step. The payoff thereby selected is the maximum out of the maximum payoffs i.e., maximax.
c) Identify the strategy corresponding to the maximax payoff.

Test Yourself 5

Pretty, a cosmetics dealer wants to decide on the amount of advertising necessary for her products. The following is the profit table corresponding to the advertisement expenditure:

Demand	Low advertisement	Average advertisement	High advertisement
High	155	165	185
Average	140	150	160
Low	130	120	125

Which strategy should Pretty adopt under the maximax approach?

Answer

The payoff table

Strategies → States of nature ↓	Low advertisement	Average advertisement	High advertisement
High demand	155	165	185
Average demand	140	150	160
Low demand	130	120	125

Steps explaining calculation of maximax strategy:

Step 1

Find out the maximum payoff for each strategy:

Low advertisement	155
Average advertisement	165
High advertisement	185

Step 2

Select the maximum out of the maximum payoffs for each strategy as identified in the preceding step. The payoff thereby selected is the maximum out of the maximum payoffs i.e., the maximax payoff is185.

Step 3

The corresponding strategy is to go for high advertisement. Therefore, the maximax approach suggests that the advertisement level should be high.

---

2. Maximin

This is a pessimistic approach to decision-making. Under this approach, the alternative that maximises the minimum most value of the outcomes for every alternative strategy is selected.

Steps followed in this approach:

a) Find out the minimum payoff for each of the strategies.
b) Select the maximum out of the minimum payoffs for each strategy as identified in the preceding step. The
c) payoff thereby selected is the maximum out of the minimum payoffs i.e. maximin.
d) Identify the strategy corresponding to the maximin payoff.

Example

Continuing the above example to understand how the decision can vary under the maximin approach.

Strategies → States of nature ↓	Low advertisement	Average advertisement	High advertisement
High demand	155	165	185
Average demand	140	150	160
Low demand	130	120	125

Steps explaining calculation of the maximin strategy:

Step 1

Find out the minimum payoff for each strategy:
Low advertisement 130
Average advertisement 120
High advertisement 125

Step 2

Select the maximum out of the minimum payoffs for each strategy as identified in the preceding step. The payoff thereby selected is the maximum out of the minimum payoffs i.e. the maximin payoff is 130.

Step 3

Corresponding strategy is to go for low advertisement

Therefore the maximin approach suggests that the advertisement level should be low.

---

3. Minimax regret

This is a strategy that tries to minimise maximum possible regret (i.e. opportunity loss) by adopting a strategy. In this approach the following steps are to be adopted:

a) For each of the states of nature find out the regret.
b) Regret = Maximum payoff for a state of nature – Payoff for the strategy
c) This will give a regret table.
d) From the regret table, find out the maximum regret for each strategy.
e) Select the minimum from the maximum regret selected above – minimax.
f) Select the strategy corresponding to the minimax regret.

Example

Continuing the above example under the minimax regret approach:

Step 1

Find out the highest payoff for each state of nature:

Strategy → States of nature ↓ Demand	Low advertisement	Average advertisement	High advertisement	Highest payoff for the states of nature
High	155	165	185	185
Average	140	150	160	160
Low	130	120	125	130

Continued on the next page

Step 2

Calculate regret table

Strategy → States of nature ↓ Demand	Low advertisement	Average advertisement	High advertisement
High	(185 - 155) = 30	(185 - 165) = 20	(185 - 185) = 0
Average	(160 - 140) = 20	(160 - 150) = 10	(160 - 160) = 0
Low	(130 - 130) = 0	(130 - 120) = 10	(130 - 125) = 5

Step 3

Maximum regret for each strategy:

Low advertisement	30
Average advertisement	20
High advertisement	5

Step 4

Minimum regret from these payoffs 5

Step 5

Corresponding strategy is high advertisement.

Therefore the minimax regret approach suggests that the advertisement level should be high.

Test Yourself 7

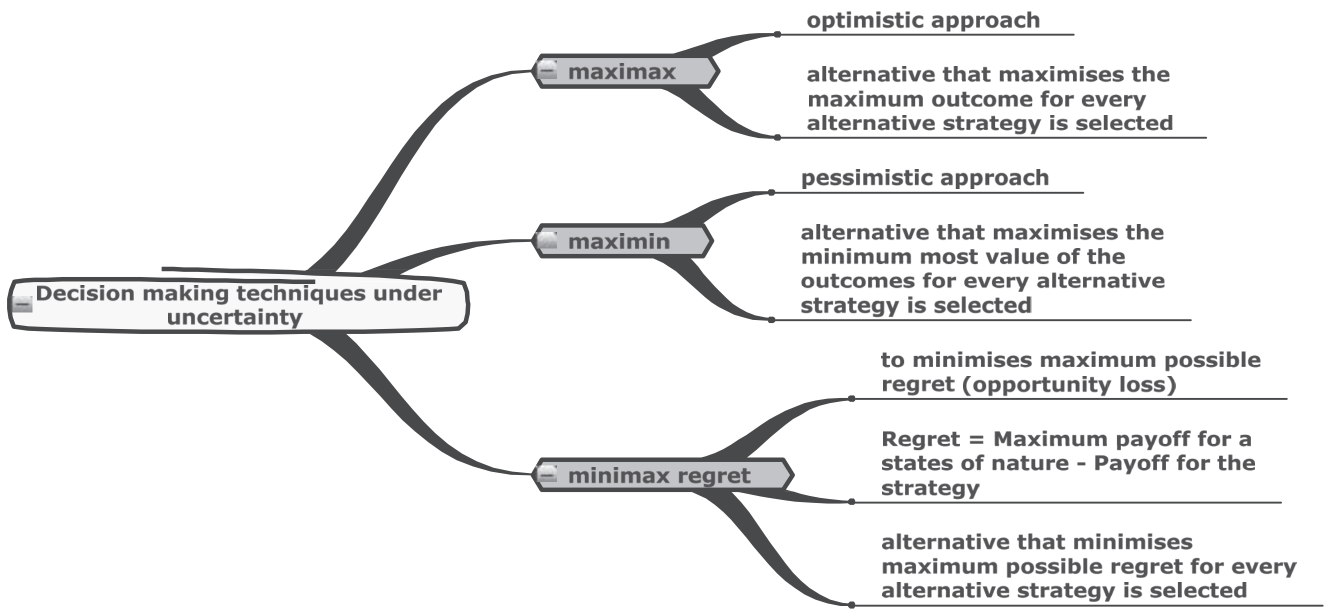
Sweetbert, a confectionery shop owner likes to buy 110 kg of either chocolate-flavoured pastry or strawberry-flavoured pastry from the producer at the beginning of the week. The chocolate-flavoured pastry costs her Tshs6,000 per kg, which she sells at Tshs8,000 per kg. Similarly, the strawberry-flavoured pastry costs her Tshs4,000 per kg which she sells at Tshs7,000 per kg. The potential demands per week for these pastries are expected to be 90, 100 or 110 kg. Any pastries remaining unsold at the end of the week are repurchased by the producer for Tshs3,000 per kg.

Required:

Identify which of the pastries Sweetie should buy, using:

(a) Maximin criterion
(b) Maximax criterion
(c) Minimax regret criterion

SUMMARY



4. The value of perfect and imperfect information

(i) Perfect information

If perfect information about the future were available, it would be very easy to make a decision as the uncertainty and the risk associated with it would be non-existent. Thus, management would like to know the cost of obtaining the perfect information about future events. This is technically known as "expected value of perfect information (EVPI)". It is the price that one would be willing to pay in order to gain access to perfect information of an uncertain outcome in decision theory.

Arithmetically, EVPI can be defined as the difference between the payoff under certainty and the payoff under risk.

EVPI = Expected value under certainty – Maximum Expected Monetary Value (EMV)

Where,

Expected value under certainty = (Best outcome or consequence for 1st state of nature) x (Probability of 1st state of nature) + (Best outcome for 2nd state of nature) x (Probability of 2nd state of nature) + (Best outcome for last state of nature) x (Probability of last state of nature)…..

Expected value under certainty is also known as expected value with perfect information (i.e. EVwPI)

Example

Strong Base Constructions is popular for its fancy construction designs. The following data relates to the estimated net profit in both favourable and unfavourable market conditions.

	Favourable market (Tshs'000)	Unfavourable market (Tshs'000)
To build a large facility	200,000	110,000
To build a small facility	175,000	55,000
Not to build any facility	-	-

The best outcome for the state of nature "favourable market" is "build a large facility" with a payoff of $200,000. The best outcome for "unfavourable" is "do nothing" with a payoff of Tshs0.

Expected value under certainty = (Tshs200,000,000) x (0.50) + (Tshs0) x (0.50)
= Tshs100,000,000

Continued on the next page

Assume that the maximum EMV is Tshs40,000,000, which is the expected outcome without perfect information.

Thus,

EVPI = EVwPI – Maximum EMV
= Tshs100,000,000 - Tshs40,000,000
= Tshs60,000,000

It means that the company would get additional benefit of Tshs60,000,000 if perfect information about the future outcome were available. The company would be ready to forgo this additional benefit by paying for such information. So the maximum cost of the perfect information will be Tshs60,000,000.

Expected Opportunity Loss (EOL)

Before we understand what EOL is, let us first look at the meaning of opportunity loss.

Opportunity Loss

a) Opportunity loss or "regret" is the loss that arises because the exact state of nature is not known at the time a decision is made.
b) The opportunity loss is computed by calculating the difference between the optimal decision for each state of nature and the other decision alternatives.

EOL is the probabilistic value criterion used to arrive at a selection of the most economically feasible option out of several alternatives available. For any alternative, EOL is the expected value of regret associated with that action.

Calculation of EOL

a) Let A_i be the i^{th} decision alternative.
b) Let $P(S_j)$ be the probability of the j^{th} state of nature.
c) Let $R(A_i,S_j)$ be the value of the regret for the combination of decision alternative A_i and state of nature S_j.
d) Let $EOL(A_i)$ be the expected opportunity loss for the decision alternative A_i.

$\boxed{EOL(A_i) = P(S_j)R(A_i,S_j)}$

Example

A food vendor has two options - whether to sell soft drinks or hot dogs in cool or warm weather. The following table shows pay-off (expected profit) of selling soft drinks and hot dogs under cool and warm weather.

	Course of action(A_j)	
	Sell soft drinks	Sell hot dogs
Events (E_i)		
Cool weather	Tshs100,000	Tshs200,000
Warm weather	Tshs400,000	Tshs250,000
Probability	0.50	0.50

Expected opportunity loss for selling soft drinks in a cool weather = Highest possible profit for an event E_i – Actual profit obtained for an action A_j
= Highest profit for selling hot dogs in a cool weather – Actual profit obtained from selling soft drinks
= (Tshs200,000 x 0.50) - (Tshs100,000 x 0.50)
= Tshs50,000

Tip

The lowest EOL is always equal to the EVPI

Example

Continuing the previous example

Expected Monetary Value

Selling soft drinks: (Tshs100,000 x 0.50) + (Tshs400,000 x 0.50) = Tshs250,000
Selling hot dogs: (Tshs200,000 x 0.50) + (Tshs250,000 x 0.50) = Tshs225,000

Highest EMV for selling soft drinks = Tshs250,000

EVwPI = Hot dog sales in cool weather + Soft drinks sales in warm weather
= (Tshs200,000 x 0.50) + (Tshs400,000 x 0.50) = Tshs300,000

EVPI = EVwPI – EMV
= Tshs300,000 - Tshs250,000
= Tshs50,000
= EOL

Example

Bags Co manufactures two types of travelling bags. However, at one point, the management of the company is faced with choosing one of the two types for manufacturing. The probability matrix after market research for the two types of bags was as follows:

Courses of action	States of nature		
	Good	Fair	Poor
Type A	75%	15%	10%
Type B	60%	30%	10%

Expected profits at different levels of market conditions are as below:

Courses of action	States of nature		
	Good (Tshs'000)	Fair (Tshs'000)	Poor (Tshs'000)
Type A	35,000	15,000	5,000
Type B	50,000	20,000	(3,000)

Required:

Calculate the expected value, expected opportunity loss and expected value with perfect information of the alternatives and advise the management of Bags Co.

Answer

Decision matrix can be set up as follows:

Courses of action	States of nature			Expected monetary value (EMV)
	Good	Fair	Poor	
Type A	75%	15%	10%	
	Tshs35,000,000	Tshs15,000,000	Tshs5,000,000	= Tshs26,250,000 + Tshs2,250,000 + Tshs500,000 = 29,000,000
Type B	60%	30%	10%	
	Tshs50,000,000	Tshs20,000,000	(Tshs3,000,000)	= Tshs30,000,000 + Tshs6,000,000 - Tshs300,000 = Tshs35,700,000

Continued on the next page

On the basis of EMV, management is advised to manufacture type B.

Expected value with perfect information

Courses of action	States of nature		
	Good	Fair	Poor
Type A	75%	15%	10%
	Tshs35,000,000	Tshs15,000,000	Tshs5,000,000
Type B	60%	30%	10%
	Tshs50,000,000	Tshs20,000,000	(Tshs3,000,000)

EVwPI = Type B in good nature + Type B in fair nature + Type A in poor nature
= Tshs30,000,000 + 6,000,000 + Tshs500,000
=Tshs36,500,000

EVPI = EVwPI – EMV
= Tshs36,500,000 - Tshs35,700,000
= Tshs800,000

EOL

Courses of action	States of nature		
	Good	Fair	Poor
	Tshs	Tshs	Tshs
Type A	75%	15%	10%
	50,000,000 - 35,000,000 = 15,000,000	20,000,000 - 15,000,000 = 5,000,000	5,000,000 - 5,000,000 = 0
Type B	60%	30%	10%
	50,000,000 - 50,000,000 = 0	20,000,000 - 20,000,000 = 0	5,000,000 - (3,000,000) = 8,000,000

EOL

Product A = (Tshs15,000,000 x 0.15 + Tshs5,000,000 x 0.15) = Tshs3,000,000
Product B = (Tshs8,000,000 x 0.10) = Tshs800,000

On the basis of EOL, management is advised to manufacture type B.

Example

Based on market research, following estimates have been established by a vendor of flowers. The cost of purchasing one flower is Tshs1,000 and the selling price per flower is Tshs1,400. Hence profit on units sold = Tshs400.

Daily demand	0	100	200	300
Probability	0.20	0.30	0.30	0.20

How many flowers can be stocked each day?
Note: assume that inventory at the end of the day cannot be sold on any other day.

Calculation of expected value (Amounts in Tshs'000)

Stock	Probability	Demand / Event of nature				EMV
(Events)		0	100	200	300	
0	0.20	0	0	0	0	0
100	0.30	(100)	40	40	40	6
200	0.30	(200)	(60)	80	80	(30)
300	0.20	(300)	(160)	(20)	120	(72)

Highest amount is Tshs6,000, hence EMV = Tshs6,000

Note: EMV column calculation is done as follows:

Continued on the next page

For third row of inventory and second column of demand

Inventory is 200 and demand is 100. Hence total cost will be 200 x Tshs1,000 = Tshs200,000 and selling price = 100 x Tshs1,400 = Tshs140,000

Profit = Sales – Cost
= Tshs140,000 - Tshs200,000
= Tshs60,000 Loss

The rest of the calculations are done in a similar way.

The optimal inventory action is the one with the highest EMV, i.e. inventory 100 flowers.

Amounts in Tshs'000

Expected value with perfect information = Maximum pay-off for each event x Probability

EVwPI = (0 x 0.20) + (40 x 0.30) + (80 x 0.30) + (120 x 0.20)
= 0 + 12 + 24 + 24
= 60
= Tshs60,000

With the existing information, the best decision that the decision maker could make was to inventory 100 flowers and earn Tshs6,000. With perfect information (forecast), the decision maker could make as much as Tshs60,000.

Hence, the expected value of perfect information is Tshs60,000 - Tshs6,000 = Tshs54,000. This is the maximum price the decision maker is willing to pay for additional information.

(ii) Imperfect information

Imperfect information refers to the information that reduces uncertainties, but, unlike perfect information does not eliminate the uncertainties.

The prime difference between perfect and imperfect information lies in the fact that the later can lead to untrue predictions. Such information is ascertained in addition to perfect information as a result of further enquiry / research.

The analysis of imperfect information is similar to that of perfect information. The expected value of such information has to be analysed before obtaining the same (in order to ascertain its worth). This is technically known as "expected value of imperfect information (EVII). It is also sometimes referred to as expected value of sample information (EVSI).

Three types of probabilities can be analysed to determine EVII:

a) Prior probabilities: these probabilities are established before the additional information is acquired

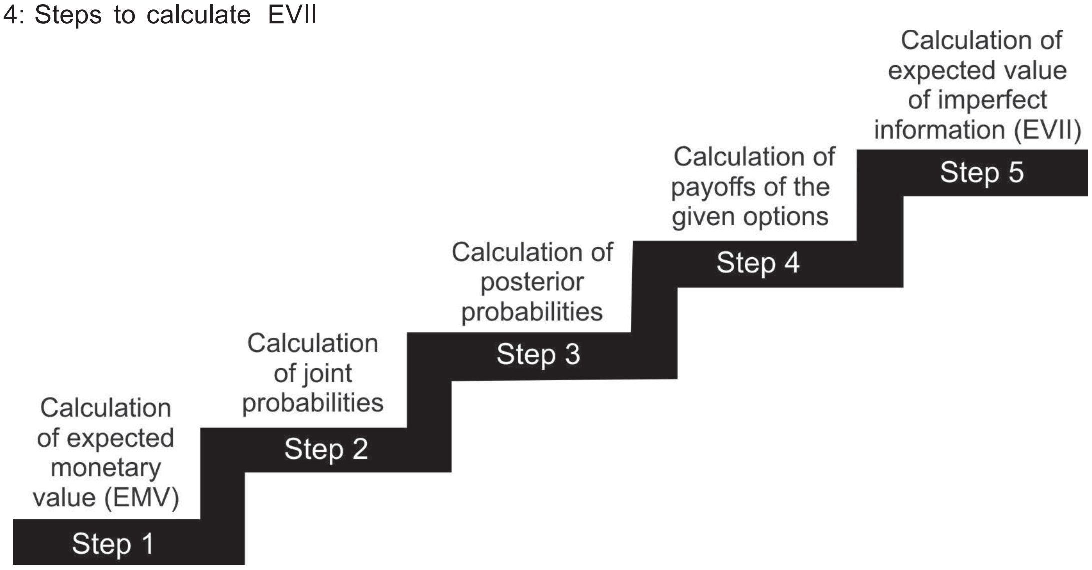
b) Conditional probabilities: refers to the probability of occurrence of an event as a result of occurrence of some other event. These are acquired in the process of acquiring the additional information.

c) Posterior probabilities: the prior probabilities need to be 'conditioned' using the conditional probabilities in order to incorporate the additional information into them. The conditioned probabilities are referred to as posterior probabilities

Steps to calculate EVII

The following diagram shows the steps involved calculating EVII

Diagram 4: Steps to calculate EVII



The process of calculating EVII is illustrated with the help of the following example.

Example

Kitchen Magic produces and sells small kitchen appliances. The marketing department of the company is considering the following options:

a) Running an end of season sales promotion (B_1)
b) Not running any sales promotions (B_2)

The following table shows the probabilities associated with three possible demand outcomes and the profit (loss) earned by the company under various options.

		Alternative options	
Possible demand situation	Probability	B_1 (Tshs'000)	B_2 (Tshs'000)
Highly favourable (D_1)	0.70	500,000	0
Favourable (D_2)	0.10	200,000	0
Unfavourable (D_3)	0.20	(300,000)	0

The marketing department of Kitchen Magic obtains additional information through a research that if it undertakes surveys, its sales can be increased. The following further options are being considered:

Option 1: Conduct survey in Northern Europe (O_1), which may help in increasing the sales by 20%
Option 2: Conduct survey in Southern Europe (O_2), which may help in increasing the sales by 15%
Option 3: Do not conduct survey at all (O_3). This will not lead to any increase in sales of the company.

Conditional probabilities associated with the survey results are given as under

Demand situation	O_1	O_2	O_3
D_1	0.60	0.30	0.10
D_2	0.30	0.60	0.10
D_3	0.10	0.10	0.80

Calculate the maximum amount that can be paid for the research.

Step 1: Calculation of expected monetary value (EMV)

		Alternative options	
Possible demand situation	Probability	B_1 (Tshs'000)	B_2 (Tshs'000)
Highly favourable (D_1)	0.70	500,000	0
Favourable (D_2)	0.10	200,000	0
Unfavourable (D_3)	0.20	(300,000)	0

Continued on the next page

EMV (B_1) = (0.7 x 500,000,000) + (0.1 x 200,000,000) + (0.2 x (300,000,000)) = Tshs310,000,000
EMV (B_2) = 0
Therefore, the maximum EMV = Tshs310,000,000

Step 2: Calculation of joint probabilities

The next step involves calculation of joint probabilities by multiplying the prior probabilities with the respective conditional probabilities.

Demand situation	Prior probabilities	O_1	Joint probabilities	O_2	Joint probabilities	O_3	Joint probabilities
D_1	0.7	0.60	0.42	0.30	0.21	0.10	0.07
D_2	0.1	0.30	0.03	0.60	0.06	0.10	0.01
D_3	0.2	0.10	0.02	0.10	0.02	0.80	0.16
Cumulative			0.47		0.29		0.24

In column four, the joint probabilities of demand situations and option1 have been computed in the following manner:

P (D_1 and O_1) = 0.7 x 0.60 = 0.42
P (D_2 and O_1) = 0.1 x 0.30 = 0.03
P (D_3 and O_1) = 0.2 x 0.10 = 0.02

Similarly, the joint probabilities of demand situations and options 2 and 3 have been computed in columns six and eight respectively.

Step 3: Calculation of posterior probabilities

The prior probabilities need to be revised.

Demand situation	Posterior probability (O_1)	Posterior probability (O_2)	Posterior probability (O_3)
D_1	0.894	0.724	0.292
D_2	0.064	0.207	0.042
D_3	0.043	0.069	0.667

The posterior probabilities can be computed with the help of the data contained in the table of step 2.

For example the posterior probabilities in column two have been computed in the following manner:

P (D_1 / O_1) = P (D_1 and O_1) / P (O_1) = 0.42 / 0.47 = 0.894

P (D_2 / O_1) = P (D_2 and O_1) / P (O_1) = 0.03 / 0.47 = 0.064

P (D_3 / O_1) = P (D_3 and O_1) / P (O_1) = 0.02 / 0.47 = 0.043

The posterior probabilities in columns three and four can be computed in a similar way.

Step 4: Calculation of payoffs of the three options

Option 1 (Conduct survey in Northern Europe): (500,000,000 x 0.894) + (200,000,000 x 0.064) + ((300,000,000) x 0.042) = Tshs 447,200,000

Option 2 (Conduct survey in Sothern Europe): (500,000,000 x 0.724) + (200,000,000 x 0.207) + ((300,000,000) x 0.069) =Tshs382,700,000

Option 3 (Do not conduct survey at all): Nil

Continued on the next page

Step 5: Calculation of expected value of imperfect information (EVII)

EVII = EMV (with survey) – Maximum EMV

EMV (with survey) = the sum of (payoffs calculated for each option (in step 4) x cumulative posterior probability of each option)

(Tshs 447,200,000 x 0.47) + (Tshs382,700,000 x 0.29) + (Tshs0 x 0.24) = 210,184,000 + 110,983,000 + 0 = Tshs 321,167,000

Therefore, EVII = Tshs321,167,000 - Tshs310,000,000 = Tshs11,167,000

Tshs11,167,000 represents the amount that Kitchen Magic would be willing to pay for the survey.

However, even after spending Tshs11,167,000, the uncertainty regarding the demand and sales would not be totally eliminated.

---

3. Construct a decision tree.
   Apply a decision tree in decision making.
   Apply the concept of decision analysis in accounting and business situation.
   [Learning Outcomes d, e and h]

In the earlier sections, we have seen that the process of decision making involves selecting an alternative that is the best fit for the organisation under the circumstances. Most complex decisions are made under risk and uncertainty. Risk is the probability of not achieving the desired objective. Uncertainty is the lack of complete knowledge about the future event, which is required for the purpose of making a decision.

A decision maker is faced with a challenge of making a decision under the circumstances of risk and uncertainty. The risk is usually measured in terms of probability or frequency and the impact or consequence. The decision maker estimates probabilities of various outcomes, based on the past information and the available knowledge about the future. The sum of all probabilities would be 1.0, so that the set of outcomes in a decision problem should be mutually exclusive and collectively exhaustive.

Decision tree is a useful analytical tool for classifying the range of alternative courses of action and their possible outcomes. It is a diagram showing several possible courses of action and possible events (i.e. states of nature) and the potential outcomes for each course of action. It depicts in a systematic manner all possible sequences of decisions and consequences. Decision trees are designed to illustrate the full range of alternatives and events that can occur under all predicted conditions. It brings out logical analysis of a problem and enables a complete strategy to be drawn to cover all eventualities before an organisation becomes committed to a decided course of action.

In the decision tree, various act-event combinations and resulting pay-offs of the problems are considered based on the risks. Decision trees give a clear, concise and meaningful overview of outcomes of various alternatives.

Decision trees provide an effective method of decision making because they:

a) clearly lay out the problem so that all options can be challenged;
b) allow us to fully analyse the possible consequences of a decision;
c) provide a framework within which to quantify the values of outcomes and the probabilities of achieving them;
d) and
e) help us to make the best decisions on the basis of existing information and best guesses.

Following criteria are used for decision making under conditions of risks and uncertainty:

(a) Expected Monetary Value (EMV)
(b) Expected Opportunity Loss (EOL)
(c) Expected Value of Perfect Information (EVPI)

Example

D Risk Limited is currently working on a process which, after paying for direct expenses, would bring a profit of Tshs12 million. The following alternatives are available to the organisation:

(i) It can conduct research (R1) which is expected to cost Tshs10 million with 90% chances of success. If it proves to be successful, it can earn a gross income of Tshs25 million.
(ii) It can conduct research (R2) which is expected to cost Tshs8 million with a chance of 60% success; the expected gross income will be Tshs25 million.
(iii) It can pay royalty of Tshs6 million for a new process which will bring a gross income of Tshs20 million.
(iv) It can continue with the current process.

Due to limitation of resources it can conduct only one of the two types of research at a time.

Required:

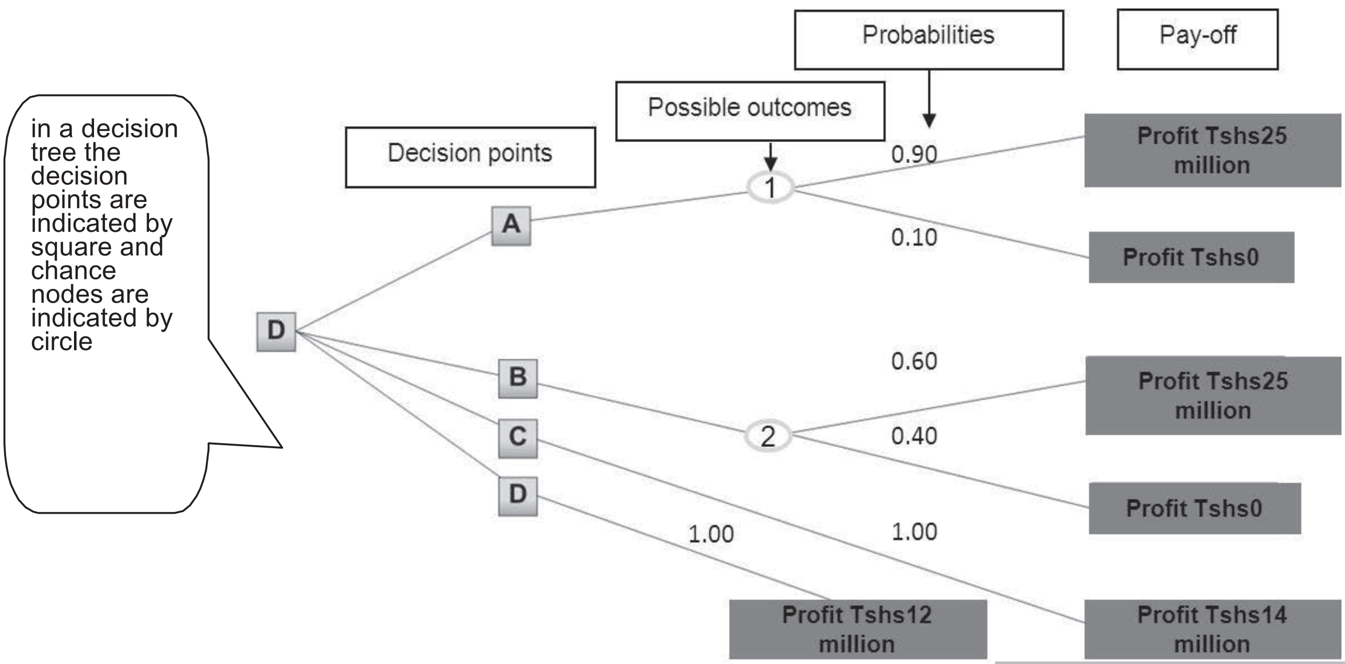
Use decision tree analysis to determine the optimal strategy for the organisation.

Solution

The various activity–event combinations and the resulting pay offs of the problem are given in the following decision tree diagram. The net expected monetary value (EMV) of each event / decision points are also indicated below:

Alternatives

A Conduct research R 1
B Conduct research R 2
C Pay royalty for new process
D Continue the current process



Continued on the next page

In the given decision tree, there are three types of nodes represented, namely decision nodes, chance nodes and terminal nodes.

Tip

Decision nodes: represent points at which the company has to make a choice of one alternative from a number of possible alternatives (e.g. points A, B, C, D mentioned above).

Chance nodes: represent points at which chance, or probability, plays a dominant role and reflect alternatives over which the company has (effectively) no control (e.g. points 1 and 2 mentioned above).

Terminal nodes: represent the ends of paths from left to right through the decision tree (e.g. pay-off).

Decision analysis at point D

Decision	Event	Probability	Gross Income(Tshs million)	Expected Income (Tshs million)
1. Conduct Research (R1)	Successful Unsuccessful	0.9 0.1	25 0	22.5 0
			Total Cost Net EMV	22.5 (10 12.5
2. Conduct Research (R2)	Successful Unsuccessful	0.6 0.4	25 0	15
			Total Cost Net EMV	15 (8) 7
3. Pay royalty for new process	Certain	1	20	20
			Total Cost Net EMV	20 (6) 14
4. Continue the current process	Certain	1	12	

As the net EMV is the highest for the alternative "Pay royalty for the new process", the optimal decision would be to introduce the new process on royalty basis.

Advantages of the decision tree method

a) Decision trees are able to generate understandable rules.
b) Decision trees perform classification without requiring much computation.
c) Decision trees are able to handle both continuous and categorical variables.
d) Decision trees provide a clear indication of which fields are most important for prediction or classification.

Disadvantages of the decision tree method

1. Decision trees are less appropriate for estimation tasks where the goal is to predict the value of a continuous attribute.
2. Decision trees are prone to errors in classification problems with many classes and relatively small number
3. of training examples.
4. The process of developing a decision tree is expensive. At each node, each candidate splitting field must be
5. sorted before its best split can be found. In some algorithms, combinations of fields are used and a search
6. must be made for optimal combining weights. Pruning algorithms can also be expensive since many
7. candidate sub-trees must be formed and compared.
8. Decision trees are not very effective in the case of non-rectangular regions. Most decision-tree algorithms
9. only examine a single field at a time. This leads to rectangular classification boxes that may not correspond
10. well with the actual distribution of records in the decision space.

4. Define central limit theorem and use it in the concept of sample means.
   Explain the appropriate sampling distributions of the sample means and sample proportions.

   [Learning Outcomes f and g]

4.1 Sampling basics

1. Population and sample

Population means the entire set of data from which a sample is selected and about which the investigator wishes to draw conclusions. The procedure of selecting a sample from the population is known as sampling.

Example

For example, a doctor examines a few drops of blood as sample and draws conclusions about the blood constitution of the whole body. In this case, the blood of the whole body is population and the drops taken for examination are the sample.

Sampling is required to save time and cost. When the data is very large, each and every member of the population is difficult to test at the time sampling is done. Sampling saves time while retaining the confidence level.

For example, if one wants to purchase food grains from a wholesaler, he might take a few grains to check the quality and then make a decision whether to purchase or not. In this case, it is not possible to check the entire warehouse of food grains to judge the quality. This will require a lot of time.

However, sampling may not always be appropriate. If a population contains a low number of high value items, and other means do not provide sufficient appropriate evidence, it may be more appropriate to carry out a 100% examination. For example, if the classroom has only 20 students and one wants to check their BMI, it is not very appropriate to do sampling of say 3 students and then take a judgement on the population.

Sampling unit: the elements of a population which are individuals to be sampled from the population and cannot be further subdivided for the purpose of the sampling at a time are called sampling units. For example, to know the average yield of maize, each farm owner's yield of maize is a sampling unit.

Sampling frame: sample frame is the list of all the items of a population.

2. Statistical and non-statistical sampling

Definition

Statistical sampling is an approach to sampling possessing the following characteristics:

(i) Random selection of the sample items; and
(ii) The use of the probability theory to evaluate sample results, including measurement of sampling risk.

A sampling approach that does not possess the above mentioned characteristics (i) and (ii) is considered non-statistical sampling.

(a) Statistical sampling

The two characteristics that are essential before a sampling can be said to be a "statistical sampling" are random selection of a sample and use of probability theory to evaluate the results.

(i) Random selection

It means that sample items are selected at random. The selection process is such that all items in the population have an equal chance of being represented in the sample. For this, the probability theory is used.

(ii) Systematic selection

It means that sample items are selected by using a constant interval between the selected items. The first item is selected on a random basis, and then, say, every 100^{th} item is selected. It should be ensured that the sample interval does not correspond to a particular pattern in the population.

(iii) Monetary unit sampling

This method of sampling aims to ascertain the accuracy of financial accounts. The sample size selection and evaluation is in monetary terms, wherein monetary amounts are converted into units. For example, a payable balance of Tshs3 million contains Tshs3 million sampling units. This method is generally used in testing internal controls.

The steps involved in sampling are as follows:

1. identifying a sample size;
2. choosing the sample;
3. carrying out audit procedures;
4. analysing the results of the audit procedure; and
5. deciding the genuineness of the population.

Use of probability theory to evaluate sample results

If a sample has been selected at random then a statistical tool of probability theory can be used to evaluate the results.

(b) Non-statistical sampling

As discussed above a sampling approach that does not have characteristics of a statistical sample is considered non-statistical sampling. The following are some methods of sampling used:

Example

The first ten invoices of each month are selected for checking. This is a judgemental or haphazard sampling and not statistical sampling in a strict sense.

(i) Haphazard selection

Under this method, sample items are selected arbitrarily, usually by manual choice. The objective of this selection process is that all items in the population have an equal chance of being represented in the sample. Therefore the selection process must be unbiased (i.e. the auditor must not deliberately avoid certain items) and free from predictability. The limitation of this method of selection is that there is no way to ensure that the estimates derived will be unbiased.

(ii) Sequence or block selection

It means that sample items are selected sequentially or in a block. E.g. 100 sales invoices (Sr No 300 to 399) are selected. This kind of sample may not be representative of the population and items may get picked from one day or period only. Therefore this method of selection is rarely used. This method of selection can be successful only if many blocks are selected.

3. Methods of selecting samples

(a) Simple random sampling

Here, the population strength is considered finite and any random samples are selected for investigation. It can be done with replacement or without replacement. Each population has equal probability of getting selected.

Example

For example, in a group of 500, one wants to know the smoking habits of the population, he might take a sample of any 50 persons. Here, those 50 are selected randomly.

In 'without replacement' method, elements of the population can enter only once. If out of 50, any one member is selected and then while selecting the second member, available population members would be 499.

However, in 'with replacement' method, population units may enter the sample more than once.

(b) Stratified random sampling

This method of sampling is widely used in surveys. The main reason behind the usage of this method is to reduce the population heterogeneity and to increase the efficiencies of the estimates.

Stratification means division into groups. In this method the population is divided into a number of subgroups or strata and then random sampling is performed.

Example

For example, a university has 2,000 students. Sample of 250 students is to be drawn and the entire population is divided into two strata. Strata A consist of 1,200 male students and Strata B consists of 800 female students. The sample using proportional allocation will be drawn as follows:

Total Population: $N = N_1 + N_2$
$N = 1,200 + 800$ where, $N_1 = 1,200$ male students and $N_2 = 800$ female students
Size of sample is 250 students.

Sample size will be proportionally derived as:
$n_1 = 250 \times 1,200/2000 = 150$
$n_2 = 250 \times 800/2,000 = 100$

(c) Systematic sampling

This is the most easy and convenient method of sampling and hence it is extensively applied in the real scenarios. Here, sampling is done on systematic basis. For example, out of the 1,000 garments, every 5th, 10th, 15th …garment is selected. It is also known as quasi random sampling. Here, sampling interval is derived using the following formula:

Sampling interval: $K = N/n$ = Population size/Sample size

4. Sampling errors and non-sampling errors

Although sampling is advantageous, there is a risk of two types of errors in a sample survey; they are sampling errors and non - sampling errors.

(a) Sampling errors: even though a sample is a part of population, it cannot be usually expected to provide full information about the population. So there is a possibility of difference between statistics and parameters in each case. The discrepancy between a parameter and its estimate due to sampling process is known as sampling error.
(b) Non-sampling errors: in all surveys some errors may occur during collection of actual information. These errors are called non-sampling errors.

The general purpose of sampling is to test a sample and draw conclusions about the population with the help of probability theory.

5. Parameters and statistics

(a) Parameters: the terms (such as mean, median, mode) describing the characteristics of a population are called parameters.

(b) Statistics: the terms (such as mean, median, mode) describing the characteristics of a sample are called statistics.

Usually, population parameters are unknown and sample statistics is used as their estimates.

Sampling distribution is the distribution of all possible values which can be assumed by some statistic measured from samples of the same size ‘ n’ randomly drawn from the same population of size N.

Sample mean

The population mean is denoted as μ and the sample mean is denoted as $\bar{X}$.
The population S.D. is denoted as a and the sample S.D. is denoted as s.

The sample mean $\bar{X}$ is defined as:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + ... + X_n}{n}$$

It is very important to note that although the mean of the distribution of $\bar{X}$ is identical to the mean of the population distribution, the variance is much smaller for large sample sizes.

Example

If the size of population is 100 and the size of its sample is 50, and on the other hand the size of population is 1,000 and the size of its sample is 100, the variance will be different in both the cases.

The mean of a representative sample provides an estimate of the unknown population mean. However, instinctively if multiple samples are taken from the same population, the estimates would vary from one another.

Distribution of the sample mean

Suppose, sampling is done over and over of the same population and mean is derived each time, it is possible to display it graphically. The graphical display of the frequency distribution of the sample means is referred to as the sampling distribution of the sample means.

Population mean: $\mu = \frac{LX}{N}$

Population standard deviation: $a = \sqrt{\frac{L(X - lJ)^2}{N}}$

Similarly, sample standard deviation is calculated as: $S = \sqrt{\frac{L(X - \bar{X})^2}{n - 1}}$ (here, n = number of elements in sample).

Standard error

The standard deviation of the sampling distribution of a statistic is known as standard error (S. E.).

S.E. of the mean = $\frac{a}{\sqrt{n}}$

S.E. gives an index of the precision of the estimate of the parameter. S.E enables us to decide the probable limits within which the population parameter may be estimated to lie.

Note: Standard Deviation defines the variability of a population or a sample. Standard Error defines the variability of an estimator that is typically a function of the whole sample.

4.2 Central limit theorem

This is one of the fundamental concepts of statistics. According to the central limit theorem, if one has a population with mean lJ and standard deviation a and he takes sufficiently large random samples from the population with replacement, then the distribution of the sample means will be nearly normally distributed.

This will hold true irrespective of whether the source population is normal or skewed, provided the sample size is adequately large (usually n > 30).

If the population is normal, then the central limit theorem holds true even for samples smaller than 30.

In fact, this also holds true even if the population is binomial, provided that min (np, n(1-p))> 5, where n is the sample size and p is the probability of success in the population.

This explains that the normal probability model can be used to quantify uncertainty when making interpretations about a population mean based on the sample mean.

The sample is considered a sampling distribution of the sample means. When all of the possible sample means are computed, then the following properties hold true:

(a) If the population has a normal distribution, then the sample means will have a normal distribution. And if the population is not normally distributed, but the sample size is sufficiently large, then the sample means will have an approximately normal distribution
(b) The mean of the sample means = The mean of the population
(c) The variance of the sample means = The variance of the population divided by the sample size
(d) The standard deviation of the sample means (known as the standard error of the mean) < The population mean
(e) The standard deviation of the sample means (known as the standard error of the mean) = The standard deviation of the population divided by the square root of the sample size

Z-score when working with the sample means: $Z = \dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

In case the sample size is more than 5% of the population size and the sampling is done without replacement, then a rectification needs to be made to the standard error of the means. For this, we need to multiply the standard error with the square root of the quotient of the difference between the population and sample sizes and deduct one from the population size.

$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}$ N is the population size and n is the sample size

Test Yourself 8

The central limit theorem states that the sampling distribution of any statistic will be normal or nearly normal:

A If the sample size is large enough
B If its sample size is small enough
C If its sample size is 10% of the population
D If its sample size is 40% of the population

Answers to Test Yourself

Answer to TY 1

The correct option is C.

Total probability is always 1

For project A:
Probability of Tshs3,000,000 = 1 – 0.6
= 0.4
For project B:
Probability of Tshs6,000,000 = 1 – (0.4 + 0.3)
= 0.3

Answer to TY 2

The correct option is A.

The product of the payoffs and probabilities gives us conditional values hence the payoff table is also called a ‘conditional value table’.

Answer to TY 3

The correct option is A.

Opportunity loss is the difference between the payoff of an act and the highest payoff corresponding to the event. For Event E_1, the highest payoff is 80 corresponding to act A_3. The opportunity loss values for E_1 will be:
For act A_1: Tshs80,000 – Tshs25,000 = Tshs55,000
For act A_2: Tshs80,000 - Tshs50,000 = Tshs30,000
For act A_3: Tshs80,000 – Tshs80,000 = Tshs0

Answer to TY 4

The correct option is A.

The alternatives under decision making are called acts. Acts are the actions that are evaluated under different prevalent conditions. The different conditions subject to which the decisions need to be taken are called events. Pay offs are the different costs attached to the various events and acts combinations. Expected values are the outcomes of any particular course of action that is followed.

Answer to TY 5

The correct option is D.

Expected profit = Tshs10 million

Profit / loss	Probability	Expected values
120 million	0.45	54 million
(80 million)	0.55	(44 million)
		10 million

Answer to TY 6

The correct option is D.

It is computed with a decision table containing conditional values for all events or states of nature. It is determined for each event or state as the sum of possible payoffs of the alternative, weighed by the probability of that payoff occurring. It measures conditional values in monetary units.

Answer to TY 7

In the given problem, the alternative strategies are:
S_1 – to buy the chocolate flavoured pastry
S_2 – to buy the strawberry flavoured pastry

The states of nature are:
N_1 – demand for 90 kg,
N_2 – demand for 100 kg,
N_3 – demand for 110 kg

Payoff (profit) can be calculated as follows:

Payoff = (Quantity sold x Profit per kg) – (Quantity unsold x Loss per kg)

Here, profit per kg of the chocolate flavoured pastry sold = Selling price – Cost
= Tshs8,000 - Tshs6,000 = Tshs2,000

Loss per kg of the chocolate flavoured pastry, if unsold = Cost - Resale price
= Tshs6,000 - Tshs3,000 = Tshs3,000

Similarly, profit per kg of the strawberry flavoured pastry sold = Tshs7,000 - Tshs4,000 = Tshs3,000
Loss per kg of the strawberry flavoured pastry, if unsold = Tshs4,000 - Tshs3,000 = Tshs1,000

Hence for the chocolate flavoured pastry (i.e. strategy S_1), as the quantity purchased is 110 kg,

Profit if the demand is 90 kg	(90 kg x Tshs2,000) – (20 kg x Tshs3,000) = Tshs120,000
Profit if the demand is 100 kg	(100 kg x Tshs2,000) – (10 kg x Tshs3,000) = Tshs170,000
Profit if the demand is 110 kg	(110 kg x Tshs2,000) – 0 = Tshs220,000

For the strawberry-flavoured pastry, (i.e. strategy S_2), as the quantity purchased is 110 kg,

Profit if the demand is 90 kg	(90 kg x Tshs3,000) – (20 kg x Tshs1,000) = Tshs250,000
Profit if the demand is 100 kg	(100 kg x Tshs3,000) – (10 kg x Tshs1,000) = Tshs290,000
Profit if the demand is 110 kg	(110 kg x Tshs3,000) – 0 = Tshs330,000

Payoff table

Strategies	States of nature			Minimum profit	Maximum profit
	N_1	N_2	N_3		
S_1	Tshs120,000	Tshs170,000	Tshs220,000	Tshs120,000	Tshs220,000
S_2	Tshs250,000	Tshs290,000	Tshs330,000	Tshs250,000	Tshs330,000

(i) Maximin criterion

Steps followed in this approach:

a) Find out a minimum payoff for each of the strategies:

Strategy	Minimum profit (Tshs'000)
S_1	120
S_2	250

b) Select the maximum out of the minimum payoffs for each strategy as identified in the preceding step. The payoff thereby selected is the maximum out of the minimum payoffs i.e. the maximin payoff is Tshs250,000.

c) Identify a strategy corresponding to the maximin payoff.

S_2 belongs to the strategy of buying the strawberry flavoured pastry, hence, it is the optimal strategy, therefore Sweetie should buy it.

(ii) Maximax criterion

Steps followed in this approach:

a) Find out a maximum payoff for each of the strategies:

Strategy	Maximum profit (Tshs'000)
S_1	220
S_2	330

b) Select the maximum out of the maximum payoffs for each strategy as identified in the preceding step. The payoff thereby selected is the maximum out of the maximum payoffs i.e. the maximax payoff is Tshs330,000.

c) Identify a strategy corresponding to the maximax payoff.

S_2 belongs to the strategy of buying the strawberry flavoured pastry, hence, it is the optimal strategy, therefore Sweetie should buy it.

(iii) Minimax regret criterion

Steps followed in this approach:

a) For each of the states of nature find out the regret

Regret = Largest payoff for a state of nature – Payoff for the strategy
We find out the regrets for each strategy, corresponding to each state of nature as:

Regret = Largest payoff for a state of nature – Payoff for the strategy

The following regret table is constructed to present the regrets for the strategies under the different states of nature.

Regret table

Strategy	N_1	N_2	N_3	Maximum regret
	(Tshs'000)			
S_1	130	120	110	130
S_2	0	0	0	0 (Minimax)

b) From the regret table, find out the maximum regret for each strategy.

Strategy	Maximum regret (Tshs'000)
S_1	130
S_2	0

c) Select the minimum of the maximum regrets selected above – i.e. the minimax regret strategy S_2 with regret of Tshs0 is selected, according to the minimax regret criteria.

d) Select the strategy corresponding to the minimax regret.

S_2 (i.e. the strategy of buying the strawberry flavoured pastry) is the optimal strategy and Sweetie should buy the strawberry-flavoured pastry.

Answer to TY 8

The correct option is A.

The central limit theorem states that the sampling distribution of any statistic will be normal or nearly normal, if the sample size is large enough.

Self Examination Questions

Question 1

A farmer is deciding whether or not to plant a crop of wheat. There are four possible outcomes of the crop:

Outcome	Profit / (loss) Tshs'000	Probability
I	5,000	25%
Ii	7,000	35%
Iii	(2,000)	10%
Iv	(500)	30%

What will be the expected value of profit?:

A Tshs3,350,000
B Tshs3,700,000
C Tshs4,000,000
D None of the above

Question 2

The Opportunity Loss table gives us the values indicating:

A The opportunity loss due to failure to take the best possible action
B The profit due to selection of the best action
C The values of best possible actions
D The loss due to failure to take action

Question 3

In a cinema hall with 500 seats, there is a 40% chance of 80% occupancy, a 10% chance of 90% occupancy and a 50% chance of 100% occupancy. Each ticket is sold at Tshs10,000. The total expected sales revenue is:

A Tshs4,550,000
B Tshs5,000,000
C Tshs4,000,000
D None of the above

Question 4

A table showing the expected payoffs along with the conditional payoffs is shown below

Pay-off table

Event	Probabilities	Conditional pay off (acts) (Tshs'000)				Expected payoff (acts) (Tshs'000) (Probability x conditional payoff)			
		A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
E_1	0.1	14	13	12	11	1.4	1.3	1.2	1.1
E_2	0.2	14	16	15	14	2.8	3.2	3.0	2.8
E_3	0.4	14	16	18	17	5.6	6.4	7.2	6.8
E_4	0.3	14	16	18	20	4.2	4.8	5.4	6.0

From the above if we have to maximise the profits (in Tshs'000) what act will you choose?

A Act with an EMV of 14.00
B Act with an EMV of 15.70
C Act with an EMV of 16.80
D Act with an EMV of 16.70

Question 5

A Regret Table or the Opportunity Loss Table for a certain events-acts combinations is shown below

Regret table

Event	Probabilities	Conditional opportunity loss (Tshs'000) Act (Purchase per day)				Expected loss (Tshs'000) Act (purchase per day)			
		A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
E_1	0.1	0	1	2	3	0	0.1	0.2	0.3
E_2	0.2	2	0	1	2	0.4	0	0.2	0.4
E_3	0.4	4	2	0	1	1.6	0.8	0	0.4
E_4	0.3	6	4	2	0	1.8	1.2	0.6	0

If you have to minimise the expected loss what act will you choose? An act with an EOL(in Tshs'000) of:

A 3.80
B 2.10
C 1.00
D 1.10

Question 6

An ice-cream vendor sells ice-creams for Tshs2,000 each. He buys 50 litres each day for the purpose of sale. There are certain days in the week when the ice-cream sells in large quantity and some days when it sells in low quantity. Based on past data, the quantity of ice-cream that can be sold and probabilities attached to it are given below:

Quantity of ice-cream (litres)	Probabilities
40	0.5
50	0.3
35	0.2

Based on the above, how many litres of ice-cream can he expect to sell?

A 7
B 20
C 15
D 42

Question 7

The expected returns and the probabilities for the shares of the three companies are given below. Suppose a customer decides to create a portfolio consisting of these three companies.

What will his expected rate of return on investment for Tshs1 million invested be?

Rate of return	Probability
15%	0.20
20%	0.60
35%	0.20

A Tshs30,000
B Tshs120,000
C Tshs70,000
D Tshs220,000

Question 8

A vendor sells roses for Tshs1,000 each. He buys 100 roses each day in order to sell them. There are certain days in the week when the flowers sell in large quantities and some days when they sell in low quantities. Based on past data, the number of flowers that can be sold and probabilities attached to it are given below.

Number of flowers	Probabilities
60	0.4
80	0.2
90	0.4

Based on the above what is the expected number of flowers that he can sell?

A 24
B 16
C 76
D 64

Question 9

Flora, a florist, sells flowers at Tshs5, 500 per bouquet. The direct cost of a bouquet is Tshs3,500. Overheads attributable to the bouquet are Tshs750. Unsold bouquets are sold at the end of the day at Tshs4,000.

Flora has estimated the following demand for a day with the corresponding probabilities:

Demand (in bunch)	0	1	2	3	4	5
Probability	0.01	0.07	0.08	0.3	0.5	0.04

Find out the optimal production of flowers.

(Assume that all unsold bouquets will be sold)

Question 10

Taba Ltd is a shoe manufacturer. In the last two years, the company has not introduced a new variety of shoes. It has Tshs100 million surplus cash for which management is exploring introducing three new product lines, each with different levels of price, profitability and sales volumes. Market research has been conducted to estimate customer demand for each type of shoe. Each type of shoe requires an initial investment of Tshs50 million, so the directors are faced with a decision to choose two out of three options.

The three new shoe products have been named TAA, BAA and DAA. TAA has been estimated as having a 90 per cent take-up from customers as it is a variation of an existing model which has been very successful. BAA has a take-up estimate of 85 per cent; however, DAA has been estimated at only 55 per cent as it is a new design that has been targeted as a diversion away from market trends, as a signature item. The Taba Ltd's marketing department has given the unit sales estimations along with the sales price set for each new product. The finance department has provided the cost estimates of each new shoe to complete the unit profit analysis.

A summary of the data is as follows:

Product	Take Up (%)	No. of Units	Sales Value (Tshs'000)	Cost (Tshs'000)	Profit (Tshs'000)
TAA	90%	10,000	395,000	158,000	237,000
BAA	85%	10,000	445,000	190,000	255,000
DAA	55%	7,500	595,000	190,000	405,000

In case the company decides not to introduce any new product, management can park the surplus cash in a liquid financial instrument and wait for alternative investment opportunities. Management is also worried that the brand of the company may lose popularity due to lack of innovativeness in introduction of new products in the market.

Required:

Construct a decision tree showing various options for enabling management to arrive at a decision.

Answers to Self Examination Questions

Answer to SEQ 1

The correct option is A.

Outcome	Profit / (loss) Tshs'000	Probability	Expected values (Tshs'000)
i	5,000	0.25	1,250
ii	7,000	0.35	2,450
iii	(2,000)	0.10	(200)
iv	(500)	0.30	(150)
		Total expected value	3,350

Answer to SEQ 2

The correct option is A.

The 'Opportunity Loss Table' contains the opportunity loss due to failure to take the best possible action. Since it is an 'Opportunity Loss Table' it never shows profits. The values of best possible actions will constitute payoffs. Loss due to failure to take action is not comparative loss and as such not a part of the 'Opportunity Loss Table'.

Answer to SEQ 3

The correct option is A.

No. of tickets sold	Sales revenue @ Tshs10,000	Probability	Expected value
80% of 500 i.e. 400	Tshs4,000,000	0.40	Tshs1,600,000
90% of 500 i.e. 450	Tshs4,500,000	0.10	Tshs450,000
100% of 500 i.e. 500	Tshs5,000,000	0.50	Tshs2,500,000
		Total expected value	Tshs4,550,000

Answer to SEQ 4

The correct option is C.

Total EMV

Act A_1 = 1.4 + 2.8 + 5.6 + 4.2 = 14.0
Act A_2 = 1.3 + 3.2 + 6.4 + 4.8 = 15.7
Act A_3 = 1.2 + 3.0 + 7.2 + 5.4 = 16.8
Act A_4 = 1.1 + 2.8 + 6.8 + 6.0 = 16.7

The sum of the pay offs for each act gives us the EMV of the chosen act. Since we want to maximise profit we choose act A_3 as it has the maximum EMV (i.e. 16.8). Thus we select this act as the optimal solution as it gives us the highest possible profits.

Answer to SEQ 5

The correct option is C.

Total EOL

A_1 = 0 + 0.4 + 1.6 + 1.8 = 3.8
A_2 = 0.1 + 0 + 0.8 + 1.2 = 2.1
A_3 = 0.2 + 0.2 + 0 + 0.6 = 1.0
A_4 = 0.3 + 0.4 + 0.4 + 0 = 1.1

The aim here is to minimise the regret or the EOL. Hence we choose the act with the minimum EOL which is to launch Invader.

The act A_3 has the minimum EOL (i.e. 1.0) we therefore select this as the optimal act. In the EOL approach we try to minimise the losses and hence select the act with minimum value of the expected loss.

Answer to SEQ 6

The correct option is D.

The expected volume of ice-cream to be sold can be calculated as below:
Sum of (Volume of ice-cream x Probability for each)

Litres of ice-cream	Probabilities	Expected values
40	0.5	40 x 0.5 = 20
50	0.3	50 x 0.3 = 15
35	0.2	35 x 0.2 = 7
Total		42

The other options give the expected sales quantities of x at the different levels of sales volumes.

Answer to SEQ 7

The correct option is D.

The expected return on investment will be calculated as the sum of the return for each security multiplied by the probability.

Rate of return	Probability	Return in Tshs'000	Expected return Tshs'000
15%	0.20	1,000 x 15% = 150	150 x 0.20 = 30
20%	0.60	1,000 x 20% = 200	200 x 0.60 = 120
35%	0.20	1,000 x 35% = 350	350 x 0.20 = 70
Total return			220

The other options are the expected returns from each security. Option A – 30, option B – 120 and option C – 70 as given in the above table.

Answer to SEQ 8

The correct option is C.

The expected number of flowers to be sold can be calculated as below
Sum of (Number of flowers x probability for each)

Number of flowers	Probabilities	Expected values
60	0.4	60 x 0.4 = 24
80	0.2	80 x 0.2 = 16
90	0.4	90 x 0.4 = 36
Total		76

The other options give the expected sales quantities of flowers at the different levels of sales volumes.

Answer to SEQ 9

Step 1

Identify strategies and states of nature

Strategies will be producing 0, 1, 2, 3, 4, 5 bouquets of flowers.

State of nature is demand for bouquets of flowers.

Step 2

Calculate payoffs

Payoff = profit / loss on sale of bouquets of flowers.

Cost of a bouquet = Direct costs + Overheads
= Tshs3,500 + Tshs750
= Tshs4,250

Profit (when bouquets are sold) = Tshs5,500 – Tshs4,250 = Tshs1,250

Loss (when bouquets are unsold) = Tshs4,000 – Tshs4,250 = (Tshs250)

Payoff will be calculated as:

Profit x Number of bouquets of flowers sold + Loss x Number of bouquets of flowers unsold

Therefore the conditional payoff table will be – (Amounts in Tshs'000)

Demand for bouquets	Probability of demand	Strategies- Number of bouquets 0	1	2	3	4	5
0	0.01	0	(0.25)	(0.50)	(0.75)	(1)	(1.25)
1	0.07	0	1.25	1	0.75	0.50	0.25
2	0.08	0	1.25	2.50	2.25	2	1.75
3	0.30	0	1.25	2.50	3.75	3.50	3.25
4	0.50	0	1.25	2.50	3.75	5	4.75
5	0.04	0	1.25	2.50	3.75	5	6.25

Step 3

From payoff tables calculate EMV for each strategy (Amounts in Tshs'000)

Payoff table

Demand for bouquets	Probability of demand	Strategies- Number of bouquets 0	1	2	3	4	5
0	0.01	0	(0.0025)	(0.0050)	(0.0075)	(0.01)	(0.0125)
1	0.07	0	0.0875	0.07	0.0525	0.035	0.0175
2	0.08	0	0.1	0.2	0.18	0.16	0.14
3	0.30	0	0.375	0.75	1.125	1.05	0.975
4	0.50	0	0.625	1.25	1.875	2.5	2.375
5	0.04	0	0.05	0.1	0.15	0.2	0.25
EMV		0	1.235	2.365	3.375	3.935	3.745

Step 4

Find out the strategy with the maximum profit payoff

As the EMV for Strategy 4 is the highest, the daily production and sale of 4 bouquets of flowers is the most profitable strategy.

6 Samples, Estimation and Confidence

Get Through Intro

One of the important roles of statistics is to make interpretations about unidentified population parameters based on sample statistics.

There are two broad areas of statistical interpretation, namely, estimation and hypothesis testing.

This Study Guide focuses on the technique of estimation. Estimation is the statistical process of defining a probable value for a population parameter (e.g., the accurate population mean or population proportion) based on a random sample.

In practice, samples are selected from the target population and sample statistics (e.g., the sample mean or sample proportion) are used as estimates of the unknown parameter.

The selected sample should represent the population with participants selected at random from the population. It is also important to quantify the precision of estimates from different samples in generating estimates.

Learning Outcomes

a) Determine point estimations for mean, proportion and standard deviation.
b) Construct interval estimations (confidence intervals) for mean, proportion, differences of two populations and paired observations.
c) Apply the concept of estimation in accounting and business situation.

1. Determine point estimation for mean, proportion and standard deviation.
   Construct interval estimations (confidence intervals) for mean, proportion, differences of two populations and paired observations.
   Apply the concept of estimation in accounting and business situation.

[Learning Outcomes a, b and c]

In statistics, estimation is the process by which one makes interpretations about a population, based on the information obtained from a sample. It is widely used by statisticians to estimate population parameters. As studied in the central limit theorem, sample means are used to estimate population means; and sample proportions, to estimate population proportions.

1.1 Estimate of population

An estimate of a population parameter may be stated in two ways: point estimation and interval estimation

1. Point estimation

A point estimate of a population parameter is a single value of a statistic.

The sample mean $\overline{X}$ is a point estimate of the population mean μ. Similarly, the sample proportion p is a point estimate of the population proportion P.

2. Interval estimation

An interval estimate is defined by two numbers, between which a population parameter is said to lie.

For example, a < x < b is an interval estimate of the population mean μ. It indicates that the population mean is greater than 'a' but less than 'b'.

1.2 Confidence Intervals

Confidence intervals are used by statisticians to express the precision and uncertainty associated with a particular sampling method. It consists of three portions; confidence level, statistic and a margin of error.

The confidence level defines the uncertainty of a sampling method. The statistic and the margin of error define an interval estimate that describes the precision of the sampling method.

The interval estimate of a confidence interval is defined by the sample statistic ± margin of error.

Example

A statistician computes an interval estimate of a population parameter and defines interval estimate as a 93% confidence interval. This means that if the same sampling method is applied to select variety of samples and calculate different interval estimates, the real (actual) population parameter would fall within a range defined by the sample statistic ± margin of error 93% of the time.

Tip

Since confidence intervals show the precision and the uncertainty of the estimate, they are ideal for point estimates.

(a) Confidence Level

Confidence level is defined as the probability portion of a confidence interval. It describes the chance that a particular sampling method will produce a confidence interval that includes the real population parameter.

Interpretation of confidence level

Example

A school wants to survey the number of students reading the newspaper daily. It has collected all possible samples from a given population, and calculated confidence intervals for each sample.

Some of the confidence intervals would include the correct population parameter; while others would not.

A 98% confidence level means that 98% of the intervals contain the exact population parameter; a 92% confidence level means that 92% of the intervals contain the population parameter; and so on.

(b) Margin of Error

In a confidence interval, the range of values above and below the sample statistic is called the margin of error.

Example

Continuing the above example of school survey,

The survey analysis reports that the students read the newspaper on a daily basis 50% of the given time period assuming that the survey had a 5% margin of error and a confidence level of 94%.

These findings result in the following confidence interval: The school is 94% confident that the students will read the newspaper daily between 45% and 55% of the given time period.

Important

In practice, many public opinion surveys report interval estimates, but not confidence intervals. They simply provide the margin of error, but not the confidence level.

To clearly interpret survey results, as a student, you need to know both!

Test Yourself 1

Identify the incorrect statement with regards to statistical estimation.

A When the margin of error is small, the confidence level is high and when the margin of error is small, the confidence level is low.
B A confidence interval is a type of point estimate and a population mean is an example of a point estimate.
C Both A and B
D Neither A nor B

Standard error of the sample mean

Recap of the concept of normal distribution

Assume that our sample has come from a normally distributed population. For any normal distribution, we can easily determine what proportions of observations in the population occur within certain distances from the mean:

50% of population falls between lJ ± 0.674a
95% of population falls between lJ ± 1.960a
99% of population falls between lJ ± 2.576a

By assuming the sample mean has a normal distribution, the variance and standard deviation of the sample mean can be calculated.

The expected value of the standard deviation of the standard mean is as follows: $a_y = \frac{a}{\sqrt{n}}$

Where,

a_y = standard deviation of the standard mean
a = standard deviation of the original population
n = size of the samples

The standard deviation of the sample mean is called the standard error of the mean: $s_y = \frac{s}{\sqrt{n}}$

Where,

s_y = standard deviation of the sample mean
s = sample estimate of the standard deviation of the original population
n = size of the samples

Point estimate of population mean

Example

A survey of the weight of 150 healthy adults having a height of 6 inches was conducted. The following data were analysed:

Sample mean: 75 kg
Sample standard deviation: 3 kg

The point estimate of population mean can be considered to be 75 kg. This is because the sample mean $\overline{X}$ is the best point estimation of population mean.

Margin of error

Formula: $E = Z_{a/2} \frac{a}{\sqrt{n}}$

Where,

$Z_{a/2}$ = Z score based on the desired confidence level
a = standard deviation of population
n = size of the sample

Example

Continuing the above example

The margin of error E and the 95% confidence interval for μ can be calculated as follows:

Here, n = 150, $\overline{X}$ = 75, s = 3, a = 5% = 0.05, a/2 = 0.025

Margin of error $E = Z_{a/2} \frac{a}{\sqrt{n}} = 1.96 \times \frac{3}{\sqrt{150}} = 0.48$

95% confidence interval for $\mu = \overline{X} - E < \mu < \overline{X} + E$
$= (75 - 0.48) < \mu < (75 + 0.48)$
$= 74.52 < \mu < 75.48$

Example

A random sample of the weight of 4 members is taken from the population containing 100 members. The sample observations are: 45, 54, 67 and 58.

The estimate of population mean is derived as follows:

$$\hat{\mu} = \bar{x} = \frac{\sum x}{n} = \frac{45 + 54 + 67 + 58}{4} = 56$$

The estimate of standard error of sample mean (assuming selection of sample with replacement) is derived as follows:

$$\hat{SE}_{\bar{x}} = \frac{S}{\sqrt{n}} \cdot \frac{\sqrt{n}}{\sqrt{n-1}} = \frac{S}{\sqrt{n-1}} = \frac{62.5}{\sqrt{4-1}} = 36.13$$

$$\text{Where, } S = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = \frac{12794}{4} - \left(\frac{224}{4}\right)^2 = 3198.5 - 3136 = 62.5$$

x	x^2
45	2025
54	2916
67	4489
58	3364
$\sum x = 224$	$\sum x^2 = 12794$

In case, the selection of sample is done without replacement, the estimate of standard error of sample mean is derived as follows:

$$\hat{SE}_{\bar{x}} = \frac{S}{\sqrt{n-1}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{62.5}{\sqrt{4-1}} \cdot \sqrt{\frac{100-4}{100-1}} = 36.13 \times 0.9847 = 35.58$$

Confidence Interval Estimates for Smaller Samples

In the case where sample sizes are smaller ($n < 30$) the Central Limit Theorem does not apply, and another distribution called the t distribution must be used.

1. t-statistic definition

Assume $x_1, x_2, x_3, \ldots x_n$ is a random sample of size 'n' from a normal population with mean μ and variance σ^2,.In this case, t-statistic is defined as follows:

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

Where,

$\bar{x}$ = sample mean
μ = population mean

S = unbiased estimate of the population variance = $\sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}$

2. Properties of t-statistic

1. Similar to normal distribution, t-distribution ranges from $-\infty$ to ∞.
2. Like the normal distribution, t-distribution has a mean zero and is symmetric.
3. As the sample size increases from 30, the t-distribution, inclines towards the Normal distribution.
4. Compared to the standard normal distribution, t-distribution has a greater dispersion.

There is a different t-distribution for each sample size. We specify a particular t-distribution by giving its degrees of freedom. The degrees of freedom for the one-sample t-statistic come from the sample standard errors in the denominator of t. Since s has n-1 degrees of freedom, the t - distribution has n-1 degrees of freedom.

Tip

Confidence interval can be derived as follows in both the situations, where n > 30 and n < 30:

For n > 30;

Confidence interval: $\bar{x} \pm Z\frac{s}{\sqrt{n}}$

For n < 30;

Confidence interval: $\bar{x} \pm t\frac{s}{\sqrt{n}}$

Confidence interval for the population proportion

Formula of confidence interval for a population proportion is as follows: $\hat{p} \pm z\sqrt{\frac{pq}{n}}$

Where, n is the size of sample and Z is the appropriate value from the standard normal distribution for the desired confidence level.

Usually, the following Z values are used frequently.

Confidence level	Z-score value
80%	1.28
90%	1.645 (by convention)
95%	1.96
98%	2.33
99%	2.58

Example

A random sample of 100 books was taken from the book shop out of which 53 books were of fiction stories. Estimate the percentage of the times (with 95% confidence) you're expected to get a fiction story book at a certain time.

Since confidence level is 95%, Z = 1.96
p and q will be 0.53 and 0.47

Continued on the next page

Confidence interval for population proportion

$$= \hat{p} \pm z\sqrt{\frac{pq}{n}}$$

$$= 0.53 \pm (1.96)\sqrt{\frac{0.53\,(0.47)}{100}}$$

$$= 0.53 - (1.96)\sqrt{\frac{0.53\,(0.47)}{100}}\,,\; 0.53 + (1.96)\sqrt{\frac{0.53\,(0.47)}{100}}$$

= 0.53 – 0.098, 0.53 + 0.098
= 0.432, 0.628

To interpret the above calculated results within the context of the given scenario, it can be said that with 95% confidence the percentage of the times one should expect to get a fiction story book at certain point of time is approximately between 43% and 63%, based on the provided sample.

Confidence intervals to estimate a difference between population mean

We have already studied confidence intervals for one mean in both cases, where n > 30 and when n < 30. In this section, we shall study confidence intervals for a difference between two means when data is paired.

In many situations, it may be required to compare two groups with respect to their mean scores on a continuous outcome.

Example

A company might be interested in comparing mean diabetes levels in men and women employees, or perhaps compare blood pressure level in smokers and non-smokers. Both situations encompass comparisons between two independent groups, assuming that there are different people in the groups being compared.

Sample sizes are denoted by n_1 and n_2, means are denoted $\bar{x}_1$ and $\bar{x}_2$, and standard deviations are denoted by s_1 and s_2 in each sample.

The parameter of interest is the difference in population means, $\mu_1 - \mu_2$ in the two independent samples presentation with a continuous outcome.

The point estimate for the difference in population means is the difference in sample means.

Hence, $\mu_1 - \mu_2 = \bar{x}_1 - \bar{x}_2$,

The confidence interval is computed using either the Z or t distribution for the selected confidence level and the standard error of the point estimate.

Standard error (SE) of the difference in sample means is the pooled (combined) estimate of the common standard deviation. Assuming that the variances in the populations are similar, it is calculated as the weighted average of the standard deviations in the samples. The formula is given below:

$$\text{Standard error (SE)} = Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Where,

Sp is the pooled estimate of the common standard deviation and is calculated as:

$$Sp = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Confidence intervals for $(IJ_1 - IJ_2)$ if n_1 and n_2 are greater than 30:

$$\overline{X_1 - X_2} \pm Z.Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Confidence intervals for $(IJ_1 - IJ_2)$ if n_1 and n_2 are smaller than 30:

$$\overline{X_1 - X_2} \pm t.Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Example

Consider the following data of an educational institute:

	Boys			Girls		
Characteristic	N	$\bar{x}$	s	n	$\bar{x}$	s
Weight (kg)	10	62	2	8	54	1.8
Height (cm)	12	168	10	10	154	5

Compare mean weight in boys and girls using 95% confidence interval.

In this case, the sample is small (< 30 for both boys and girls), so we can use the confidence interval formula with t distribution.

Next, we will check the assumption of equality of population variances.

Confidence interval formula: $\overline{X_1 - X_2} \pm t.Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

Confidence interval = $(62 – 54) – t\ (1.91)\sqrt{\frac{1}{10} + \frac{1}{8}}$ and $(62 – 54) + t\ (1.91)\sqrt{\frac{1}{10} + \frac{1}{8}}$

Confidence interval = [8 – (2.12) (1.91) (0.47)], [8 + (2.12) (1.91) (0.47)]

Confidence interval = (8 – 1.91, 8 + 1.91)

Confidence interval = 6.09, 9.91 i.e. 6, 10

Refer to the t-table in the appendix of the book and use the df = 10 + 8 - 2 = 16th row. For 95% confidence level, the value of t is 2.12

Where,

$$Sp = \sqrt{\frac{(n_1 - 1)s1^2 + (n_2 - 1)s2^2}{n_1 + n_2 - 2}}$$

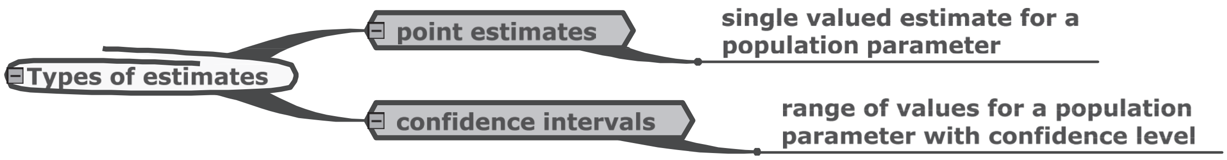
$$Sp = \sqrt{\frac{(10 - 1)4 + (8 - 1)3.24}{10 + 8 - 2}}$$

$$Sp = \sqrt{\frac{36 + 22.68}{16}}$$

Sp = 1.915

We are 95% confident that the difference in mean weight between boys and girls is between 6 and 10 kg. The best estimate of the difference, the point estimate, is 8 kg. The standard error of the difference is 0..898, and the margin of error is 1.915 kg.

SUMMARY



Test Yourself 2

In a sample of 400 professors, 184 expressed dissatisfaction regarding new syllabus given by the institute. Calculate point estimate of the population of total professors who would be dissatisfied and give an estimation of the standard error of the estimate.

Test Yourself 3

Refer to the previous example of the educational institute and compare the mean height of the boys and the girls using 95% confidence interval.

Application of estimation theory

In statistics, the process of using sample data to draw inferences about the population is known as estimation technique. Nowadays, estimating techniques are used by many industries to calculate:

(a) the cost of buying materials
(b) the cost of acquiring machineries
(c) the cost involved in the production of goods, etc.

Estimation process gives companies a rational indication of the amount to be invested.

It can also be used in the areas of auditing to assess the degree of risk in the financial statements. The theory of estimates also helps an organisation in materials management. What amount of goods need to be stocked and within what ranges can be estimated using this theory.

It is also applied in weather forecasting e.g. using point estimation, a forecast of tomorrow's weather would be made as 32 degree Celsius or using interval estimate, a forecast of tomorrow's weather would be ranging from 27 degree to 38 degree Celsius.

Answers to Test Yourself

Answer to TY 1

The correct option is C.

All of the given statements are incorrect.

Option A is incorrect because the margin of error does not affect the confidence level. When the margin of error is small, the confidence level can be either low or high or anything in between.

Option B is incorrect because a confidence interval is a type of interval estimate and a sample mean is an example of a point estimate.

Answer to TY 2

Point estimation of the population proportion: p= 184/400 = 0.46

Standard error of the proportion: $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.46 \times 0.54}{400}} = 0.025$

Answer to TY 3

In this case, the sample is small (< 30 for both boys and girls), so we can use the confidence interval formula with t distribution.

Next, we will check the assumption of equality of population variances.

Confidence interval formula: $\overline{X_1 - X_2} \pm t.Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

Confidence interval = (168 – 154) – t (8.14) $\sqrt{\frac{1}{12} + \frac{1}{10}}$ and (168 – 154) + t (8.14) $\sqrt{\frac{1}{12} + \frac{1}{10}}$

Confidence interval = [14 – (2.086) (8.14) (0.43)], [14 + (2.086) (8.14) (0.43)]
Confidence interval = (14 – 7.3, 14 + 7.3)
Confidence interval = 6.7, 21.3 i.e. 7, 21

Refer to the t-table in the appendix of the book and use the df = 12 + 10 - 2 = 20th row. For 95% confidence level, the value of t is 2.086

Where,

$$Sp = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$Sp = \sqrt{\frac{(12 - 1)100 + (10 - 1)25}{12 + 10 - 2}}$$

$$Sp = \sqrt{\frac{1100 + 225}{20}}$$

Sp = 8.14

We are 95% confident that the difference in the mean weight between boys and girls is between 7 and 21 cm. The best estimate of the difference, the point estimate, is 14 cm. The standard error of the difference is 3.5 cm, and the margin of error is 8.14 cm.

Self Examination Questions

Question 1

A sample of 144 dining table sets showed an average weight of 16 kg with a standard deviation of 1.2 kg.

Required:

(a) Construct a 97% confidence interval for the mean of the population.
(b) Construct a 68.36% confidence interval for the mean of the population.
(c) Discuss why the answers to part (a) and part (b) are different.

Question 2

A manufacturing unit of a textile industry produces 60,000 cotton shirts on a daily basis. A random sample of 1% of the shirts was taken and out of them 3% was found defective.

Required:

Estimate the number of cotton shirts that can be expected to be defective in the daily manufacturing process, assuming 95% level of confidence.

Question 3

Using α = 0.05, a confidence interval for a population proportion is determined to be 0.68 to 0.78. If the level of significance is decreased, what will be the impact on the interval for the population proportion?

A It becomes narrower
B It becomes wider
C It does not change
D It remains the same

Question 4

What is the use of an interval estimate in a range of values?

A To estimate the shape of the population's distribution
B To estimate the sampling distribution
C To estimate a sample statistic
D To estimate a population parameter

Answers to Self Examination Questions

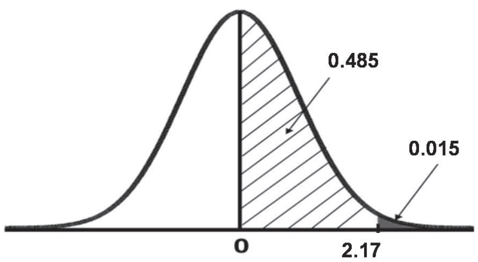
Answer to SEQ 1

Part (a)

In the given example, σ = 1.2, $\underline{x}$ = 16; n = 144, confidence level = 0.97

1 – a = 0.97 hence, a = 0.03 and a/2 = 0.015

According to the normal curve, $Z_{0.015}$ = 2.17



Since X follows normal distribution, and σ is known, the confidence intervals will be determined as follows:

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 16 \pm 2.17 \frac{1.2}{\sqrt{144}}$$

Confidence interval [(16 – 2.17(0.1), (16 + 2.17 (0.1)]

= (16 - 0.217, 16 + 0.217)

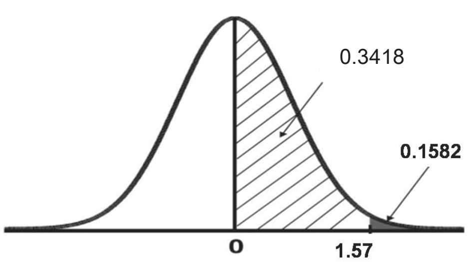
= 15.783, 16.217

Part (b)

In the given example, σ = 1.2, $\underline{x}$ = 16; n = 144, confidence level = 0.6836

1 – a = 0.6836 hence, a = 0.3164 and a/2 = 0.1582

According to the normal curve, $Z_{0.1582}$ = 1.005



Since X follows normal distribution, and σ is known, the confidence intervals will be determined as follows:

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 16 \pm 1.005 \frac{1.2}{\sqrt{144}}$$

Confidence interval [(16 – 1.57(0.1), (16 + 1.57 (0.1)]

= (16 - 0.157, 16 + 0.157)

= 15.843, 16.157

Part (c)

The answers to part (a) and part (b) are different because as the level of confidence increases, the confidence interval becomes wider. In part (a) level of confidence was 97% which decreased to 68.36% in part (b). Hence, the confidence interval also changed from (15.783, 16.217) to (15.843, 16.157).

Answer to SEQ 2

In the question, sample size n = 600 (1 % of 60,000) and p = 0.03 is given. (Hence, q will be 0.97)
Population size: N = 60,000

Hence, SE (p) will be computed as follows:

$$\sqrt{\frac{pq}{n}} . \sqrt{\frac{N - n}{N - 1}}$$

$$= \sqrt{\frac{0.03 \text{ x } 0.97}{600}} . \sqrt{\frac{60{,}000 - 600}{60{,}000 - 1}}$$

= 0.00696 x 0.99499
= 0.00693

Refer to the tip given in LO 1 for 1.96

Determining 95% confidence limit to p:
= [p – 1.96 (0.00693)], [p + 1.96 (0.00693)]
= (0.03 – 0.0136, 0.03 + 0.0136)
= 0.0164, 0.0436

The number of cotton shirts that can be reasonably expected to be defective in the daily production process at 95% of the confidence level will be computed as follows:

= (60,000 shirts x 0.0164), (60,000 shirts x 0.0436)
= 984 shirts to 2,616 shirts
Can be written as (984, 2,616)

Answer to SEQ 3

The correct option is B.

Answer to SEQ 4

The correct option is D.

An interval estimate is a range of values used to estimate a population parameter.

7 HYPOTHESIS TESTING

Get Through Intro

Hypothesis testing is a topic considered to be the heart of statistics. In statistical territory, this technique belongs to inferential statistics.

In practice, researchers are engaged in various studies of marketing, production, pharmaceutics, health science, etc. and are drawing conclusions based on the results of their studies.

However, it is imperative to know the validity of the research, its claims and its conclusions.! The ultimate goal of research is to determine the validity of these claims.

The main role of hypothesis testing is to test the accuracy of population on the basis of sample analysis. Samples are derived by using carefully and cautiously designed statistical experiments.

This Study Guide will enable you to test for mean, proportion, differences of two means, differences of two proportions, paired observations, etc. It also focuses on Karl Pearson's non-parametric test, the Chi-square.

Learning Outcomes

a) State the steps of conducting a test for mean, proportion, differences of two means, differences of two proportion and paired observations.
b) Describe the errors of decisions Type I and II errors.
c) Conduct test for mean, proportion, differences of two means, differences of two proportions and paired observations.
d) Explain the use of variance ratio test ANOVA.
e) Conduct non-parametric tests – Chi-Square for goodness and for independence.
f) Conduct rank and product moment correlation coefficient tests.
g) Apply tests of hypothesis in accounting and business situations.

1. State the steps of conducting a test for mean, proportion, differences of two means, differences of two proportion and paired observations.
   Describe the errors of decisions Type I and II errors.
   Conduct test for mean, proportion, differences of two means, differences of two proportions and paired observations.

[Learning Outcomes a, b and c]

Hypothesis testing is the method in which samples are selected to learn more about characteristics in a given population. It is a perfect and systematic approach way to test claims or ideas regarding a group or population.

1. Steps of hypothesis testing

(a) Identify hypothesis to be tested i.e. null hypothesis or alternative hypothesis.
(b) Select the criteria of testing.
(c) Select a random sample from the population.
(d) Compare what is observed in the sample to what is expected to be observed if the claim being tested is true. Here, if the discrepancy is too large, then one will likely decide to reject the claim as being not true.

Example

A newspaper published an article stating that kids under the age of 12 years in Tanzania study for an average of 3 hours a day. To test whether this claim is true, one should record the actual time, of randomly selected 10 kids, spent in studies in a day.

The mean calculated for these randomly selected kids will be considered as sample mean. It can then be compared with the population mean stated in the article.

Let us revise the chapter of estimation with the following example.

Example

The noise levels in a school have been measured in decibels. The mean and the standard deviation of the noise level in 84 corridors were recorded as 61.2 decibels and 7.9 decibels respectively. The 95% confidence interval of the actual mean will be computed as follows:

Here, n = 84, $\bar{X}$ = 61.2, a = 7.9, confidence interval = 95% = 0.95

a = 0.05; hence, a/2 = 0.05/2 = 0.025

Now, $1 - a/2 \rightarrow Z_{\frac{a}{2}} = 1.96$

$= 1 - 0.025 \rightarrow Z_{\frac{a}{2}} = 1.96$

$= 0.975 \rightarrow Z_{\frac{a}{2}} = 1.96$

Confidence intervals

$$\bar{X} - \left(Z_{\frac{a}{2}} \frac{a}{\sqrt{n}}\right) < \mu < \bar{X} + \left(Z_{\frac{a}{2}} \frac{a}{\sqrt{n}}\right)$$

$$61.2 - \left(1.96 \times \frac{7.9}{\sqrt{84}}\right) < \mu < 61.2 + \left(1.96 \times \frac{7.9}{\sqrt{84}}\right)$$

61.2 – 1.69 < μ < 61.2 + 1.69

59.51< μ < 62.9

---

2. Types of hypothesis

Null and alternative hypothesis; alternative hypothesis can be one and two tailed alternative hypothesis

Null hypothesis and alternative hypothesis

Null hypothesis is a starting point. Null hypothesis is denoted by H_0. It is indicated as the hypothesis / claim that are initially assumed to be true. Alternative hypothesis is denoted by H_1 or H_A. This is the hypothesis or claim that is initially assumed to be false but which may be decided to be accepted if there is sufficient evidence.

The Null Hypothesis is only rejected in favour of the Alternative Hypothesis if there is sufficient evidence of this “beyond reasonable doubt”.

Null hypothesis contains equalities e.g. =, $\leq$, or $\geq$ (is equal to, greater than or equal to, lesser than or equal to).

Alternative hypothesis is regarded as the complement of null hypothesis.

Example

If one wants to test the null hypothesis that the population has a specified mean, μ0 then:

Step 1: Null hypothesis H0: μ = μ0

Step 2: Alternative hypothesis may be any of the following:

(i) H1 : μ is not equal to μ0 (i.e. μ > μ0 or μ < μ0)
(ii) H1 : μ is greater than μ0 (i.e. μ > μ0)
(iii) H1: μ is smaller than μ0 (i.e. μ < μ0)

In case (i), it is a two-tailed alternative hypothesis
In case (ii), it is a right tailed hypothesis
In case (iii), it is a left tailed hypothesis

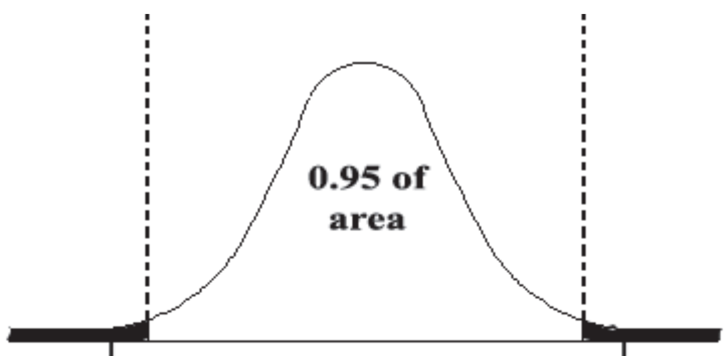
Tip

The locations of alternative hypothesis is very important since it assists us to select whether we have to use a single – tailed (right or left) or two tailed test.

3. Level of significance

As studied in the previous Study Guide 6, while testing a given hypothesis, the maximum probability with which one would be willing to take risk is called level of significance of the test. It is denoted by a. The measurement of level of significance commonly engaged in testing of significance are 0.05 and 0.01.

The following standard normal curve illustrates the area in which the null hypothesis could be accepted or rejected when it is being tested at a level of significance 0.05 and a two-tailed test is engaged.



In case the sample statistics falls in this area, the null hypothesis is accepted.

Some important definitions:

(a) Test statistic: is referring to a function of a sample of observations which provides a basis for testing the validity of the null hypothesis. Used for testing validity of the null hypothesis by referring to the function of a sample of observations.

(b) Critical region refers to the region of normal curve in which the null hypothesis is rejected when a calculated value of the test statistic lies within this region.

(c) Critical value is the value determining the boundary of the critical region.

(d) Significance level (the probability of Type I error alpha) is the size of the critical region.

(e) One tailed test: a type of test wherein the critical region is located wholly at one end of the sampling distribution of the test statistic; H_1 involves < or > but not both.

(f) Two tailed test: a type of test wherein the critical region comprises of areas at both ends of the sampling distribution of the test statistic; H_1 involves f.

4. Use of p-values

P-value is the probability of obtaining a test statistic. It is the conversion of test statistic into conditional probability. The p values are used to get the answer to the following statement: “Considering the null hypothesis to be true, determine the probability of observing the current data or data that is more extreme.”

Tip

Useful tips for p-values
When p value > 0.10 � the observed difference is “not significant”
When p value $ 0.10 � the observed difference is “marginally significant”
When p value $ 0.05 � the observed difference is “significant”
When p value $ 0.01 � the observed difference is “highly significant”
Note: here, significant means the observed difference is not likely just because of chance.

5. Type I and Type II errors

In hypothesis testing, the type of errors can be classified as Type I errors and Type II errors.

a) A Type I error occurs if an investigator rejects a null hypothesis that is actually true in the population. It is also known as false-positive.
b) A Type II error occurs if the investigator fails to reject a null hypothesis that is actually false in the population. It is also known as false-negative.

Even though Type I and Type II errors can never be avoided completely, the investigator can reduce their probability by increasing the sample size. This is because the larger the sample size, the lesser is the probability that it will differ significantly from the population.

The probability of Type I error is called the level of significance of hypothesis test and is denoted by α (alpha) and the probability of Type II error is denoted by β (beta).

Diagram 1: Type I error and Type II error

Statistical decision	Null hypothesis H_0	
	H_0 True	H_0 False
Reject H_0	Type I error (α)	Correct
Accept H_0	Correct	Type II error (β)

- type I error: null rejected when it is actually true
- type II error: null failed to rejected when it is actually false
- power = 1 – β (the probability of rejecting a false H_0)

Probabilities of Type I error and Type II error are related with each other. However, in order to reduce the probability of Type II error, rather than increasing the probability of Type I error, it is advisable to increase the sample size.

Tip

The probability of Type I error is generally chosen as 1%, 5% or 10%.

Generally, Type I errors are considered the more serious and hence in hypothesis testing procedure the probability of these errors (a) are controlled by the investigators and they are usually unaware of the probability of Type II errors (◊).

6. Steps of significance test

(a) Conversion of research equation into null or alternative hypothesis
(b) Calculation of test statistic of data
(c) P value and conclusion
(d) Decision (if P $ a, the null hypothesis will be rejected. Otherwise it will be retained for want of evidence)

1.1 Test concerning number of success

Example

In a fashion show, 520 females and 480 males participated for a ramp walk. Does this situation confirm that the females and males – participated in equal number?

Answer

H0 = Females and males participated in equal number
H1 = Females and males participated in unequal number
Here, p = q = 0.5

Sample size n = 1,000

Since it follows binomial probability distribution, S.E. will be derived as: $\sqrt{npq} = \sqrt{1000 \times 0.5 \times 0.5} = 15.81$

Difference between observed and estimated male / female = 520 – 500 and 480 – 500 = 20 and (20)

$$Z = \frac{\text{Difference}}{\text{S.E.}}$$

For female: Z = 20/15.81 = 1.27

Since the computed amount 1.27 is less than the 1.96 S.E. (at 5% significance level), null hypothesis is accepted, that is, female and male members participated in equal number

Test Yourself 1

750 citizens of Tanzania from a specific region were randomly selected. Out of them, 420 citizens were lefty and the rest were righty. Can it be assumed that the region comprise of lefty and righty people in equal proportion?

1.2 Test concerning proportion of success

In this type of scenario, instead of recording the number of success as calculated above, the proportion of success in each sample shall be recorded.

Mean proportion of success: $p = \frac{\text{number of success in the sample}}{\text{sample size}}$

Standard deviation of proportion of successes: S.D. = $\sqrt{\frac{pq}{n}}$ and hence standard error (S.E.) of the proportion os success will be $\sqrt{\frac{pq}{n}}$

Example

1,000 ball pens were randomly selected from a large consignment and out of those selected, 130 ball pens were found to be defective.

Here,

Mean proportion of success: $p = \frac{\text{number of success in the sample}}{\text{sample size}}$

$= \frac{\text{number of defective pens}}{\text{total pens selected as sample}}$

$= \frac{130}{1,000}$

= 0.13

Hence, q = 0.87

Standard error = $\sqrt{\frac{pq}{n}} = \sqrt{\frac{0.13\,(0.87)}{1000}}$ = 0.0106

Certain limit within which % of defective pens lies can be calculated as follows:

$p \pm 3$ S.E.
= 0.13 – 3 (0.0106), 0.13 + 3 (0.0106)
= 0.13 – 0.0318, 0.13 + 0.0318
= 0.0982, 0.1618
= 9.82% to 16.18%

1.3 Test concerning difference between proportions

As studied in the previous Study Guide 10, sometimes two or more samples are drawn from two or more different populations. Here, it may be either interesting or necessary to know if the difference between the proportion of successes is significant or not.

Here, hypothesis will be the existence of difference between p_1 and p_2 because of fluctuations of random sampling.

Standard error (p_1 - p_2) = $\sqrt{\frac{pq}{n_1}} + \sqrt{\frac{pq}{n_2}}$

Where, p is considered as pooled (combined) estimate of the actual population proportion.

$$p = \frac{n_1p_1 + n_2p_2}{n_1 + n_2}$$

$$Z = \frac{p_1 - p_2}{S.E._{(p1-p2)}}$$

Tip

In case Z is less than 1.96 S.E. at 5% significance level, the difference is regarded as not significant.

Example

1,000 people were randomly selected to study their smoking habits. Out of them, 800 people were found to be regular smokers. Later on, excise duty increased on cigarettes. At that time, 800 people were found to be regular smokers out of 1200 sample size.

There has been a significant reduction in the smokers due to rise in excise duty. Is this statement true?

Answer

Here, two population studies have been carried out:

(i) Before rise in excise duty
(ii) After rise in excise duty

Let n_1 = 1,000 and n_2 = 1,200

Hence, p_1 = 800/1000 i.e. 0.8 and p_2 = 800/1200 i.e. 0.67

Pooled (combined) estimate of the actual population proportion

$$p = \frac{n_1p_1 + n_2p_2}{n_1 + n_2}$$

$$p = \frac{800 + 800}{1000 + 1200} = 0.73 \text{ and hence } q = 0.27$$

Standard error (p_1 - p_2)

$$= \sqrt{\frac{0.73\,(0.27)}{1,000}} + \sqrt{\frac{0.73\,(0.27)}{1,200}}$$

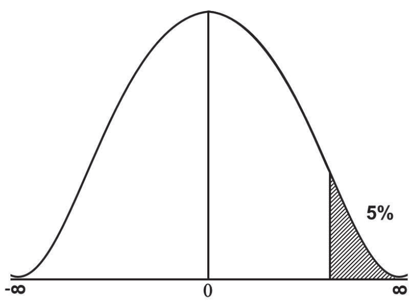
= 0.014 + 0.013
= 0.027

In this case,

Null hypothesis: H0: $p_1 \leq p_2$; there is no significant reduction in the smokers due to rise in excise duty
Alternative hypothesis: H1: $p_1 > p_2$; there is significant reduction in the smokers due to rise in excise duty

$$\text{Statistical test: } Z = \frac{\text{Difference}}{\text{S.E.}} = \frac{0.80 - 0.67}{0.027} = 4.82$$

Critical region: in this scenario, 5% of significance level gives the critical value of Z as 1.645 for a right tailed test as shown in the diagram below:



Continued on the next page

Decision: the computed statistical test value of 4.82 is greater than estimated Z value of 1.645. Hence, the null hypothesis can be rejected and it can be concluded that there is a significant reduction in the number of smokers after the rise in excise duty on cigarettes.

Test Yourself 2

250 people were randomly selected from the city of Arusha. Out of them, 100 found to be consumers of bath soap over bath gel. Another sample of 200 people was randomly selected from the city of Dodoma. Out of them, 100 were found to be consumers of bath soap over bath gel.

Is this data pointing towards a significant difference between Arusha and Dodoma when it comes to the usage of bath soap over bath gel? (assume 1% of significance level)

1.4 Test of significance for a single mean from large samples

Similar to sampling studies, in hypothesis studies also, mean and standard deviation are obtained to get an idea about population distribution. If a sample consists of more than 30 units, it is considered as large sample size and if a sample consists less than or equal to 30 units, it is considered to be a small sample size.

Steps for conducting test of significance for a single mean from large samples:

1. Set up null and alternative hypothesis
2. Compute test statistic Z as $\frac{(\bar{x}) - \mu}{S.E.(\bar{x})} = \frac{[(\bar{x}) - \mu](\sqrt{n})}{s}$
3. Set the level of significance a
4. Derive the critical value from Z score
5. Decide whether to accept or reject null hypothesis

Example

A vendor claims that he can pack 1 kg sugar at the rate of 60 packs per hour. He was asked to take 50 minutes in which he could pack 58 packs of sugar with standard deviation of 7.5 packs. Assume 5% level of significance.

Answer

Null hypothesis: H0: grocery vendor's claim is true
Alternative hypothesis: H1: grocery vendor's claim is not true

Computation of test statistic

$$Z = \frac{[(\bar{x}) - \mu](\sqrt{n})}{s} = \frac{58 - 60}{7.5}\sqrt{50} = 1.886$$ (negative is ignored as it is considered in modulus)

Level of significance is 0.05 hence critical value will be 1.96

Since, calculated value is less than the estimated value, null hypothesis is accepted. The vendor's claim of packing 60 packs per hour is true.

1.5 Test of significance of difference between two means from large samples

When two samples are derived from two populations, the role of hypothesis testing is to:

a) Test the equality of population means (whether μ1 = μ2)

b) Test the significance of the difference between two independent sample means ($\overline{x_1 - x_2}$)

Steps for conducting test of significance of difference between two means from large samples:

1. Set null and alternative hypothesis as follows

   H0: μ1 = μ2
   H1: μ1 < μ2 or μ1 > μ2 (two tailed test)

2. Compute test statistic

   When S.D. of population are known: $Z = \dfrac{\bar{x}1 - \bar{x}2}{S.E.(\bar{x}1 - \bar{x}2)}$ where, S. E = $\sqrt{\dfrac{a_1^2}{n_1} + \dfrac{a_2^2}{n_2}}$

   When S.D. of population are not known: $Z = \dfrac{\bar{x}1 - \bar{x}2}{S.E.(\bar{x}1 - \bar{x}2)}$ where, S. E = $\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

   Other steps remain the same as done before.

Example

A kitchen accessories supplier wishes to determine whether his designs made by two different designers vary or not. The following data is collected on the random sampling of two designers' designs.

	Designer 1	Designer 2
$\bar{x}$	253.4	258.7
s	32.4	33.8
n	200	100

Using the data provided by kitchen accessories supplier, test the null hypothesis that the average number of design is same for the designer 1 and designer 2. Assume normal distribution and 5% significance level.

Answer

Null hypothesis: H0: the average number of design is same for the designer 1 and designer 2 i.e. μ1 = μ2
Alternative hypothesis: H1: μ1 < μ2 or μ1 > μ2

Computation of test statistic:

When S.D. of population are not known: $Z = \dfrac{\bar{x}1 - \bar{x}2}{S.E.(\bar{x}1 - \bar{x}2)} = \dfrac{253.4 - 258.7}{4.08} = -1.30$ i.e. 1.30

Where, S. E = $\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}} = \sqrt{\dfrac{1049.76}{200} + \dfrac{1142.44}{100}} = \sqrt{5.2488 + 11.4244} = 4.08$

Level of significance: 0.05 hence critical value of Z score is 1.96

Since the computed test statistic value is less than the critical value, null hypothesis is accepted and can be concluded that the average number of design is same for the designer 1 and designer 2.

1.6 Test of significance for difference of two standard deviations – large samples

When two different standard deviations are derived from two populations, null hypothesis that $\sigma_1 = \sigma_2$ can be tested as follows:

Standard error of standard deviation = $\dfrac{a}{\sqrt{2n}}$

When the population standard deviations are known:

S.E. = $\sqrt{\dfrac{a_1^2}{2n_1} + \dfrac{a_2^2}{2n_2}}$

It is also a case of two tailed test.

When the population standard deviations are not known:

$$S.E. = \sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}$$

The above is also a case of two tailed test.

Example

A customer purchases two types of shirts from his regular suppliers A and B. The following is summarised to know whether there is difference between the standard deviation significant at 10% level of significance.

	Supplier A	Supplier B
Number of shirts purchased	100	140
Average life of shirt (in years) (till it becomes scrap)	1.4	1.5
Standard deviation (in years)	0.56	0.45

Answer

Null hypothesis: H0: the standard deviation of the two suppliers do not vary significantly

Alternative hypothesis: H1: the standard deviation of the two suppliers vary significantly (two tailed test)

Computation of test statistic

When S.D. of population are not known: $Z = \frac{\bar{x}1 - \bar{x}2}{S.E.(\bar{x}1 - \bar{x}2)} = \frac{1.4 - 1.5}{0.049} = -2.04$ i.e. 2.04

Where, $S.E = \sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}} = \sqrt{\frac{0.3136}{200} + \frac{0.2025}{240}} = \sqrt{0.0016 + 0.0008} = 0.049$

Level of significance: 0.10 hence critical value of Z score is 1.645

Since the computed test statistic is greater than critical value, reject null hypothesis. It can be concluded that the standard deviation of the two suppliers vary significantly.

1.7 Test of significance for single proportion – large samples

Steps for hypothesis testing:

Set up null hypothesis and alternative hypothesis; H0: $P = P_0$ (P_0 is a particular value of proportion); H1: $P = P_0$ (two tailed test)

Compute test statistic as follows:

Standard error of proportion = S.E. (p) = $\sqrt{\frac{PQ}{n}}$ where, n is sample size and p is population proportion

$Z = \frac{p - P}{S.E.(p)}$ where p = sample proportion

Other steps remain as it is!

Example

Random samples of 200 electric bulbs were taken out of which 6 bulbs were defective. Considering level of significance at 5%, state whether the given evidence is sufficient for concluding that production process results in more than 2% defective bulbs.

Answer

This is one tailed test as P > 0.02

Step 1:

Null hypothesis: H0: production process is in control, defects are less than or equal to 2%
Alternative hypothesis: H1: production process results in more than 2% defective bulbs

Step 2:

Test statistic: $Z = \frac{p - P}{S.E.(p)} = \frac{0.03 - 0.02}{0.0098} = 1.02$

Where, = S.E. (p) = $\sqrt{\frac{pq}{n}} = \sqrt{\frac{0.02(0.98)}{200}} = 0.0098$

Step 3: level of significance is 0.05

Step 4: critical value of Z score is 1.645

Step 5: Decision: since computed test statistic value is less than the critical value, null hypothesis is accepted and it can be concluded that production process is in control, defects are less than or equal to 2%.

1.8 Test of significance of difference between two sample proportions – large samples

Steps for hypothesis testing:

Set up null and alternative hypothesis
H0: two samples are drawn from the same population; $P_1 = P_2$
H1: P_1 is not equal to P_2 (two tailed test)

Compute test statistic

$Z = \frac{P_1 - P_2}{S.E.(P_1 - P_2)}$ where, P_1 and P_2 are proportion of population

S.E. (P1 – P2) = $\sqrt{\frac{P_1 Q_1}{n_1}} + \sqrt{\frac{P_2 Q_2}{n_2}}$

Other steps remain as it is

Test Yourself 3

In a toothpaste manufacturing company, production processes are carried out in batches. On an average, there are 5% defective toothpastes in a batch of 500. Some of the machineries were upgraded, and then, out of a batch production of 360 toothpastes, only 12 were defective.

Required:

State whether the machineries have been improved or not.

1.9 Test of significance of a single mean – small samples (where $n \leq 30$)

Steps of hypothesis testing to test significance of a single mean when sample size is small.

1. Set null and alternative hypothesis
   H0: there is no significant difference between sample mean and population mean
   H1: there is significant difference between sample mean and population mean

2. Compute test statistic

   Test statistic $t = \dfrac{\bar{x} - \mu}{S.E.(\bar{x})}$

   Where, standard error of the mean S.E. (x bar) = $\dfrac{s}{\sqrt{n-1}}$

3. Decide level of significance

4. Derive critical value from t table for (n – 1) d.f.

5. Decision: whether to accept or reject null hypothesis

Example

A cold drinks manufacturing company has its several branches across Tanzania. The average sale of a particular branch A is 70 dozen bottles. The company invested huge marketing efforts to boost sales. Afterwards, it randomly selected 26 of its branches and calculated average sales of those branches as 73.5 dozen bottles. The standard deviation was 8 bottles.

Test whether the marketing efforts are effective at 5% level of significance.

Answer

Null hypothesis: H_0: there is no significant effectiveness of marketing efforts on sales: $\mu = 70$

Alternative hypothesis: H_1: there is significant effectiveness of marketing efforts on sales: $\mu > 70$ (single tailed test)

Test statistic $t = \dfrac{\bar{x} - \mu}{S.E.(\bar{x})} = \dfrac{73.5 - 70}{1.6} = 2.19$

Where, standard error of the mean S.E. (x bar) = $\dfrac{s}{\sqrt{n-1}} = \dfrac{8}{\sqrt{26-1}} = 1.6$

Critical value of t 0.05 significance level for 25 d.f. is 1.708 for single tailed test.

Since test statistic value is greater than critical value, null hypothesis is rejected and it can be concluded that there is significant effectiveness of marketing efforts on sales.

1.10 For difference of means, paired t-test

Here, it is assumed that sample observations are not completely independent.

Steps to calculate paired t test hypothesis:

1. Set null and alternative hypothesis

   H0: $\mu 1 = \mu 2$
   H1: $\mu 1 < \mu 2$ or $\mu 1 > \mu 2$ (two tailed test)

2. Compute test statistic

   Test statistic $t = \dfrac{\overline{D}}{S.E.(\overline{D})}$

   Where,

   $\overline{D}$ = mean of difference between each pair observations

   Estimated SE of difference: $SE = \dfrac{S}{\sqrt{n}}$

   Where, $S = \sqrt{\dfrac{n\Sigma D^2 - (\Sigma D)^2}{n(n-1)}}$

3. Level of significance is usually 0.05, 0.1 or 0.01 in this case.
4. Critical value is taken from t table at the significance level for (n – 1) d.f.
5. Decision: if test statistic is less than critical value, null hypothesis is selected.

Example

A health check-up of five randomly selected kids of same age group was done (before and after Vitamin D medication) and the following results were achieved:

Kids	A	B	C	D	E
Weight in kg (before Vitamin D medication)	22	24	24.6	26.4	25
Weight in kg (after Vitamin D medication)	24	23.6	25	27.2	24.2

Conduct a test and check whether there is any difference in the weight after vitamin D medication. Consider level of significance 1%.

Answer

Step 1:

Null hypothesis: there is no change in weight after vitamin D medication i.e. $\mu 1 = \mu 2$
Alternative hypothesis: there is change in weight after vitamin D medication i.e. $\mu 1 < \mu 2$ or $\mu 1 > \mu 2$ (two tailed test)

Step 2: computation of test statistic t

Test statistic $t = \dfrac{\overline{D}}{S.E.(\overline{D})} = 0.4/0.489 = 0.82$

Where,

$\overline{D}$ = mean of difference between each pair observations = $\Sigma D/n = 2/5 = 0.4$

Kids	Weight in kg (before Vitamin D medication)	Weight in kg (after Vitamin D medication)	D	D square
A	22.0	24.0	2.0	4
B	24.0	23.6	-0.4	0.16
C	24.6	25.0	0.4	0.16
D	26.4	27.2	0.8	0.64
E	25.0	24.2	-0.8	0.64
			$\Sigma D = 2.0$	$\Sigma D^2 = 5.6$

Continued on the next page

Estimated SE of difference: SE = $\frac{S}{\sqrt{n}} = \frac{1.095}{\sqrt{5}} = 0.489$

Where, S = $\sqrt{\frac{n\Sigma D^2 - (\Sigma D)^2}{n(n-1)}} = \sqrt{\frac{5\,(5.6) - (4)}{5(4)}} = 1.095$

Step 3

Level of significance is 0.01 in this case.

Step 4

Critical value is taken from t table at the significance level for (n – 1) d.f. which is 4.6

Step 5

Decision: since test statistic is less than critical value, null hypothesis is selected and it is concluded that there is no change in weight after vitamin D medication.

1.11 Test of significance of difference between two means – small samples (where n $\leq$ 30)

Example

Tanzan Ltd has two branches, Tan and Zan. They sell heavy machineries used in the railways. The company conducted sample survey of both the branches and obtained following results: (amounts in Tshs'000)

Tan:
Sales volume 9
Average sales revenue 102.5
Standard deviation 12.5

Zan:
Sales volume 10
Average sales revenue 85
Standard deviation 10

Is there any significant difference in the mean sales of two branches Tan and Zan?

Answer:

Null hypothesis: H_0: there is no significant difference between the sales of two branches i.e. μ1 = μ2

Alternative hypothesis: H1: there is significant difference between the sales of two branches i.e. μ1 < μ2 OR μ1 > μ2 (two tailed test)

Test statistic: t = $\frac{\bar{x}1 - \bar{x}2}{S.E.(\bar{x}1 - \bar{x}2)} = \frac{102.5 - 85}{5.47} = 3.2$

Where, standard error of difference: S. E = S x $\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ = 11.897 x $\sqrt{\frac{1}{9} + \frac{1}{10}}$ = 5.47

Continued on the next page

Where, Estimated standard deviation: S = $\sqrt{\dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$ = $\sqrt{\dfrac{(9)(156.25) + (10)(100)}{9 + 10 - 2}}$ = 11.897

$$\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}} = \sqrt{\frac{0.3136}{200} + \frac{0.2025}{240}} = \sqrt{0.0016 + 0.0008} = 0.049$$

Level of significance is 0.05 and the critical value of t at a = 0.05 for (n1 + n2 – 2) 17 degrees of freedom for two tailed test is 1.74

Since the computed test statistic value is greater than the critical value of t, null hypothesis is rejected and it can be concluded that there is significant difference between the sales of the two branches.

---

2. Explain the use of variance ratio test ANOVA.

[Learning Outcome d]

Definition

According to R.A. Fisher, Analysis of Variance (ANOVA) is the “Separation of Variance ascribable to one group of causes from the variance ascribable to other group”.

Analysis of Variance (ANOVA) is a statistical method which is widely used to test differences between two or more means. The calculations are very similar to the examples given in the previous Learning Outcome. We will discuss variance ratio test ANOVA in brief.

Example

Tan Co has upgraded his machinery and now it can be used in three different production processes. Three production processes A, B and C are tested to check whether their outputs are equivalent.

The following observations of number of units produced are made during the test:

Production processes	Number of units produced							
A	15.0	18.0	19.5	16.5	15.0	21.0	22.5	19.5
B	13.5	16.5	15.0	18.0	19.5			
C	16.5	15.0	22.5	21.0	18.0	19.5		

You are required to carry out the analysis of variance and state your conclusion.

Answer

Step 1: To carry out the analysis of variance, we need to prepare some tabular calculations:

									Total	Squares
A	15.0	18.0	19.5	16.5	15.0	21.0	22.5	19.5	147.0	21,609
B	13.5	16.5	15.0	18.0	19.5	0	0	0	82.5	6,806.25
C	16.5	15.0	22.5	21.0	18.0	19.5	0	0	112.5	12,656.25
									342.0	41,071.5

Here, 342 is the grand total and 41,071.5 is considered as Total sum of squares (TSS)

Continued on the next page

Step 2: sum of squares of all the class totals (SST)

A	225.00	324.00	380.25	272.25	225.00	441.00	506.25	380.25
B	182.25	272.25	225.00	324.00	380.25			
C	272.25	225.00	506.25	441.00	324.00	380.25		

SST = 6,286.5

Test Procedure:

Null Hypothesis: H0: $\mu 1 = \mu 2 = \mu 3$; i.e., there is no significant difference between the three production processes.

Alternative Hypothesis H1: $\mu 1 \neq \mu 2 \neq \mu 3$ i.e. there is significant difference between the three production processes

Level of significance: Let α : 0.05

Test statistic

Correct factor (C.F.) = $\frac{G^2}{N} = \frac{(342)^2}{19} = 6156$

TSS (Total sum of squares) = 6286.5 – 6156 = 130.5

SST (Sum of squares due to production process) is calculated as follows:

$$\frac{21{,}609}{8} + \frac{6{,}806.25}{5} + \frac{12{,}656.25}{6} - 6{,}156$$

= 2,701.13 + 1,361.25 + 2,109.37 – 6,156

= 15.75

SSE (Sum of squares due to error) = TSS – SST = 130.5 – 15.75 = 114.75

ANOVA Table

Sources of variation	Degree of freedom	S.S	M.S.S	F ratio
Between Processes	3 - 1 = 2	15.75	=15.75/2 = 7.875	7.875/8.16 = 0.965
Error (18 – 2 = 16)	16	130.50	= 130.50/16 = 8.16	
Total	19 1 = 18			

Table Value: Table value of F_e for (2,16) degree of freedom at 0.05 level of significance is 3.63

Inference: Since calculated value of F_0 is less than table value of Fe, we may accept null hypothesis and conclude that there is no significant difference between the three production processes.

3. Conduct non-parametric tests – Chi-Square for goodness and for independence.
[Learning Outcome e]

Let us first have a brief look on parametric and non-parametric tests.

(a) Parametric Tests: Rely on theoretical distributions of the test statistic under the null hypothesis and assumptions about the distribution of the sample data (i.e., normality)
(b) Non-Parametric Tests: Referred to as "Distribution Free" as they do not assume that data are drawn from any particular distribution

A statistical test in which no hypothesis is made about specific values of parameters is regarded as non-parametric test.

One of the simplest and most extensively used non-parametric tests is Chi-Square test, first used by Karl Pearson. Symbolically, it is denoted as: χ^2 and used to designate the degree of the discrepancy between theory and observation.

$$\chi^2 = \sum_{i=1}^{n} \frac{(Oi - Ei)^2}{Ei}$$

Where,
Oi = observed frequencies
Ei = estimated frequencies

Interpretation of χ^2

In case, $\chi^2 = 0$, the observed frequencies and estimated frequencies correspond with each other. The greater the value of χ^2, the greater the degree of the discrepancy between observed and estimated frequencies are.

χ^2 Tests can be applied when the following conditions are satisfied:

(a) Each sample observations must be independent of each other.
(b) The total frequency should be greater than 50.
(c) There should not be any theoretical cell frequency of less than 5 and in case of less than 5 frequencies they
are combined together to make either 5 or more than5.
(d) χ2 test is wholly dependent on degrees of freedom.

3.1 Testing the goodness of fit

Karl Pearson proved that the statistic $\chi^2 = \sum_{i=1}^{n} \frac{(Oi - Ei)^2}{Ei}$ follows χ2-distribution under the null hypothesis, which assumes that there is no substantial difference between the observed and the theoretical values.

Where,

n = n – k – 1 degree of freedom
01, 02, ...0n are the observed frequencies,
E1 , E2…En corresponding to the expected frequencies
k = the number of parameters to be estimated from the given data.

Example

Three unbiased coins are simultaneously tossed. The number of heads occurring at each toss was noted. This was repeated 100 times with the following results.

Number of heads	0	1	2	3
Number of toss	24	30	21	25

Fit a Binomial distribution assuming under the hypothesis that the coins are unbiased.

Answer

Null Hypothesis: H0: The given data fits the Binomial distribution. i.e. the coins are unbiased.
$p = q = 0.5$; $n = 3$; $N = 100$

Computation of expected frequencies:

Number of heads	Probability: $P(X = x) = {}^{3}C_{x}(0.5)^{x}(0.5)^{3-x}$	Expected frequency N.P(X)
0	${}^{3}C_{0}(0.5)^{0}(0.5)^{3} = 0.125$	12.50
1	${}^{3}C_{1}(0.5)^{1}(0.5)^{2} = 0.375$	37.50
2	${}^{3}C_{2}(0.5)^{2}(0.5)^{1} = 0.375$	37.50
3	${}^{3}C_{3}(0.5)^{3}(0.5)^{1} = 0.125$	12.50
Total	1.000	100.00

Computation of chi square values

Observed frequencies O	Expected frequencies E	(O – E)	$(O – E)^2$	$\frac{(O-E)^2}{E}$
24.00	12.50	11.50	132.25	10.58
30.00	37.50	(7.50)	56.25	1.50
21.00	37.50	(16.50)	272.25	7.26
25.00	12.50	12.50	156.25	12.50
100.00	100.00	0.00	617.00	31.84

$$X_o^2 = \sum_{i=1}^{n} \frac{(Oi-Ei)^2}{Ei} = 31.84$$

Here, k = 0 since no parameter is estimated from the data

Expected value:

$$X_e^2 = \sum_{i=1}^{n} \frac{(Oi-Ei)^2}{Ei}$$ follows chi-square distribution with degree of freedom (n – k – 1)

i.e.
v = 4 – 0 – 1 = 3 d.f. 7.81474

Refer to the appendix: percentage points of chi-square distribution: 3rd row column 0.05

Interpretation: Since $\chi_0^2 < \chi_e^2$, null hypothesis at 5% level of significance is accepted and the given data fits Binomial distribution.

Test Yourself 4

The following data represents the distribution of number of typo errors done by a typist.

Number of errors	0	1	2	3	4
Number of pages	47	48	54	31	20

Fit Poisson distribution and test the goodness of fit.

3.2 Test of independence

Chi-square is widely applied to test the independence of attributes when the sample data is presented in the form of contingency table. Here, H0 refers to the attributes that are independent and H1 refers to the attributes which are not independent.

Example

500 employees in an IT industry were randomly selected and graded in accordance with their intelligence level and their financial status. With the help of $\chi 2$ test, determine whether there is any association between financial conditions and intelligence level.

	Financial conditions		
Intelligence level	Poor	Rich	Total
Low	125	150	275
High	175	50	225
	300	200	500

Answer

Null hypothesis: H0: there is no association between the attributes, financial condition and intelligence level
Alternative hypothesis: H1: an association between the attributes, financial condition and intelligence level exists

Calculation of test statistic:

$$E_{ij} = \frac{\text{Total of row x Total of column}}{\text{Sample size}}$$

Observed value	Expected values	O – E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
125	(275 x 300)/500 = 165	(40)	1,600	9.697
175	(225 x 300)/500 = 135	40	1,600	11.852
150	(275 x 200)/500 = 110	40	1,600	14.545
50	(225 x 200)/500 = 90	(40)	1,600	17.778
				53.872

$$\chi_o^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 53.872$$

Degree of freedom: v

= (r – 1) (c – 1)
= (2 – 1) (2 – 1)
= 1

When v = 1, $\chi 2$ for 0.05 is 3.84

The calculated value of $\chi 2$ 53.872 is greater than the expected value of $\chi 2$ 3.84. Hence, null hypothesis is rejected and alternative hypothesis is accepted that there is an association between the attributes, financial condition and intelligence level.

Test Yourself 5

400 taxi drivers were randomly selected by a transport agency and their performance were rated in accordance with their length of experience (more than a year or less than a year of driving experience) and their performance in the driving test conducted.

The data was summarised in the following table:

	Performance		Total
	Good	Poor	
Experienced	200	100	300
Not experienced	40	60	100
Total	240	160	400

Using chi square test the independence at 5% level of significance and draw your conclusion.

---

4. Conduct rank and product moment correlation coefficient tests.

[Learning Outcome f]

The detailed calculations of rank and product moment correlation coefficient, slope and Y-intercept are explained in the Study Guide 5 of this Study Text. Here, we shall restrict our explanation to conducting their tests for shape and test of significance.

4.1 Hypothesis testing concerning significance of correlation coefficient

Here, p is assumed as the coefficient of linear correlation in a bivariate normal population; r is correlation coefficient and n is sample size.

Null hypothesis: H0: p = 0 and alternative hypothesis: H1: p < 0 or p > 0 (two tailed test)

Standard error is r is derived as: $\sqrt{\frac{1-r^2}{n-2}}$

Test statistic: $t = \frac{r}{\text{S.E.of r}}$ which can be simplified as follows:

It follows t-distribution with degree of freedom n – 2.

$r.\sqrt{\frac{n-2}{1-r^2}}$

Example

The salaries of 18 pairs of males and females were studied and the correlation coefficient was derived as 0.5. Test the significance of correlation using t-test.

Answer

Null hypothesis: H0: p = 0

Alternative hypothesis: H1: p < 0 or p > 0

Test statistic: $r.\sqrt{\frac{n-2}{1-r^2}} = 0.5.\sqrt{\frac{18-2}{1-0.25}} = 0.5\ (4.62) = 2.31$

Degree of freedom: v = 18 – 2 = 16 at 5% of confidence level is 2.12

Here, the calculated value is greater than the table value, hence, reject the null hypothesis and accept the alternative hypothesis.

Example

A teacher wants to study the correlation between the scores achieved in the inter exam and the scores achieved in the final exam. He has taken 27 pairs of male students and female students from his classroom. Calculate the least value of r in a sample of 27 pairs from a bivariate normal population significant at 5% level.

Test statistic: $r.\sqrt{\frac{n-2}{1-r^2}} = r.\sqrt{\frac{27-2}{1-r^2}} = r.\sqrt{\frac{25}{1-r^2}}$

Degree of freedom: v = 27 – 2 = 25 at 5% of confidence level is 2.06

We can say that t-statistic > degree of freedom (to compute least value of r)

$r.\sqrt{\frac{25}{1-r^2}} > 2.06$

Taking square on both the sides

$\frac{25r^2}{1-r^2} > 4.2436$

$25r^2 > (4.2436)(1 - r^2)$
$25r^2 > (4.2436 - 4.2436r^2)$
$29.2436r^2 > 4.2436$

Taking square root on both the sides

$5.407r > 2.06$
$r > 2.06/5.407$
$r > 0.381$

The minimum value of r should be 0.381

Test Yourself 6

A political group derived correlation coefficient of two variables as 0.36. Compute the sample size that must have been considered by the group if t statistic is greater than degree of freedom 2.72.

4.2 Hypothesis testing for rank correlation

Similar to above illustrated various hypothesis test, for Spearman's test, a random sample is drawn, test statistics are computed and compared it to the critical value appropriate for the sample size, the required significance level and whether the test is 1- or 2-tail.

The null hypothesis should be expressed as there being no association between the variables.

5. Apply tests of hypothesis in accounting and business situations.

[Learning Outcome g]

The concept and techniques of hypothesis testing have been discussed in the previous learning outcomes of this Study Guide. Hypothesis testing can be applied in various accounting and business situations to study the population parameters based on the sample results. The following is an example of a situation where hypothesis testing can be applied in business decision:

Example

Robert, the manager (operations) of a soft drinks bottling plant has initiated a new age process for quality control. He strongly believes this new measure will significantly decrease the rate of rejection in the soft drinks' bottling line, and bring down the rejection rate of bottles from its current level of 25 per batch of 750 bottles.

He now wishes to pilot the new quality control procedure and establish whether the data supports his estimation. Robert can do so by processing the following questions:

a) What is the null hypothesis for this new procedure?
b) In the context of the new procedure, what would be the consequences of making a Type I error?
c) In the context of the new procedure, what would be the consequences of making a Type II error?

In case Robert is satisfied with the results, it would be financially viable to go ahead with the proposed new age process of quality control.

Answers to Test Yourself

Answer to TY 1

n = 750, p = q = 0.5

Expected frequency of lefty citizens: 750 x 0.5 = 375

Observed frequency of lefty citizens: 420

Difference = 45

Since it follows binomial probability distribution, S.E. will be derived as: $\sqrt{npq} = \sqrt{750 \times 0.5 \times 0.5} = 13.69$

$$Z = \frac{\text{Difference}}{\text{S.E.}}$$

For female: Z = 45/13.69 = 3.28

Since the computed amount 3.28 is more than the 1.96 S.E. (at 5% significance level), null hypothesis is rejected that the region comprise of lefty and righty people in equal proportion

Answer to TY 2

Null hypothesis: H0: there is no significant difference between Arusha and Dodoma when it comes to the usage of bath soap over bath gel.

Alternate hypothesis: H1: there is a significant difference between Arusha and Dodoma when it comes to the usage of bath soap over bath gel

Let n_1 = 250 and n_2 = 200
Hence, p_1 = 100/250 i.e. 0.4 and p_2 = 100/200 i.e. 0.5

Pooled (combined) estimate of the actual population proportion

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$p = \frac{100 + 100}{250 + 200} = 0.44 \text{ and hence } q = 0.56$$

Standard error (p_1 - p_2)

$$= \sqrt{\frac{0.44\,(0.56)}{250}} + \sqrt{\frac{0.44\,(0.56)}{200}}$$
= 0.0314 + 0.0351
= 0.0665

Statistical test: $Z = \frac{\text{Difference}}{\text{S.E.}} = \frac{0.40 - 0.50}{0.0665} = |1.504| = 1.504$

Critical region: in this scenario, 1% of significance level gives the critical value of Z as 2.58

Decision: the computed statistical test value of 1.504 is less than estimated Z value of 2.58, Hence, the null hypothesis can be accepted and it can be concluded that there is no significant difference between Arusha and Dodoma when it comes to the usage of bath soap over bath gel.

Answer to TY 3

Null hypothesis: H_0: there is no significant improvement in the machinery after upgrading i.e. $P_1 = P_2$

Alternative hypothesis: H_1: there is significant improvement in the machinery after upgrading i.e. $P_1 < P_2$ (one tailed test)

Test statistic:

$$Z = \frac{P_1 - P_2}{\text{S.E.}(P_1 - P_2)} = \frac{0.05 - 0.033}{0.0191} = 0.89$$

Where, P_1 and P_2 are proportion of population and are computed as: 0.05 (5% given) and 12/360 = 0.033

$$\text{S.E. (P1 – P2)} = \sqrt{\frac{P_1 Q_1}{n_1}} + \sqrt{\frac{P_2 Q_2}{n_2}} = \sqrt{\frac{(0.05)\,(0.95)}{500}} + \sqrt{\frac{(0.033)\,(0.967)}{360}} = 0.0097 + 0.0094 = 0.0191$$

Critical value of Z at significance level 0.05 is 1.645 for one tailed test.

Since, computed statistic value is less than critical value, null hypothesis is accepted and it can be concluded that there is no significant improvement in the machineries after upgrading.

Answer to TY 4

Null Hypothesis :

The given data fits the Poisson distribution.

Level of significance :

Let $\alpha = 0.05$

Firstly, let us calculate mean m:

Number of errors	Number of pages	f x n
0	47	0
1	48	48
2	54	108
3	31	93
4	20	80
	200	329

Mean m = 329/200 = 1.645 = 1.6 (approx.)

$e^{-1.6} = 0.2019$

Computation of E

Number of errors	Number of pages	Probability $P(X = x) = \frac{e^{-m}.m^x}{x!} = \frac{e^{-1.645}.(1.645)^x}{x!}$	Expected frequency N . P(X) = 200 x P(X) (figures are rounded off)
0	47	$0.2019 \frac{(1.6)^x}{x!} = 0.2019$	41
1	48	$0.2019 \frac{(1.6)^x}{x!} = 0.3230$	65
2	54	$0.2019 \frac{(1.6)^x}{x!} = 0.2584$	53
3	31	$0.2019 \frac{(1.6)^x}{x!} = 0.1378$	29
4	20	$0.2019 \frac{(1.6)^x}{x!} = 0.0551$	12

Computation of chi square values

Observed frequencies O	Expected frequencies E	(O – E)	$(O – E)^2$	$\frac{(O - E)^2}{E}$
47	41	6	36	0.88
48	65	(17)	289	4.45
54	53	1	1	0.02
31	29	2	4	0.14
20	12	8	64	5.33
200	200	0		10.82

$$\chi o^2 = \sum_{i=1}^{n} \frac{(Oi - Ei)^2}{Ei} = 10.82$$

Here, k = 0 since no parameter is estimated from the data

Expected value:

$$\chi e^2 = \sum_{i=1}^{n} \frac{(Oi - Ei)^2}{Ei}$$ follows chi-square distribution with degree of freedom (n – k – 1)

i.e.

v = 5 – 0 – 1 = 4 d.f 9.488

Refer to the appendix: percentage points of chi-square distribution: 4th row column 0.05

Interpretation: Since χ02 > χe2, null hypothesis at 5% level of significance is rejected and the given data do not fit the Poisson distribution.

Answer to TY 5

Null hypothesis: H0: there is no association between the attributes, experience and performance
Alternative hypothesis: H1: an association between the attributes, experience and performance exists

Calculation of test statistic:

$$E_{ij} = \frac{\text{Total of row x Total of column}}{\text{Sample size}}$$

Observed value	Expected values	O – E	$(O – E)^2$	$\frac{(O-E)^2}{E}$
200	(300 x 240)/400 = 180	20	400	2.22
100	(300 x 160)/400 = 120	(20)	400	3.33
40	(100 x 240)/400 = 60	(20)	400	6.67
60	(100 x 160)/400 = 40	20	400	10.00
				22.22

$$Xo^2 = \sum_{i=1}^{n} \frac{(Oi - Ei)^2}{Ei} = 22.22$$

Degree of freedom: v
= (r – 1) (c – 1)
= (2 – 1) (2 – 1)
= 1

When v = 1, χ2 for 0.05 is 3.84

The calculated value of χ2 22.22 is greater than the expected value of χ2 3.84. Hence, null hypothesis is rejected and alternative hypothesis is accepted that there is an association between the attributes experience and performance.

Answer to TY 6

Answer to TY

$$0.36\sqrt{\frac{n-2}{1-0.1296}} > 2.72$$

$$0.36 \text{ x } \frac{\sqrt{n-2}}{\sqrt{0.8704}} > 2.72$$

$\sqrt{n-2} > 2.72$ x (0.9329/0.36)

$\sqrt{n-2} > 7.04$

n - 2 > 49.68
n > 51.68 i.e. 52

Self Examination Questions

Question 1

Which of the following statements is correct?

A In a hypothesis test, null hypothesis refers to what we are trying to prove.
B In a hypothesis test, alternative hypothesis is always assumed to be true.
C The alternate hypothesis is accepted unless there is sufficient evidence to say otherwise.
D The null hypothesis is not rejected unless there is sufficient evidence to reject it.

Question 2

When does Type I error occur?

A When H_o is accepted
B When H_1 is accepted
C When H_o is wrongly rejected
D When H_1 is wrongly rejected

Question 3

Usually, the significance level of a statistical hypothesis test is a fixed probability and is chosen to be:

A 5%
B 99%
C 90%
D 50%

Question 4

All of the following statements regarding the null and alternative hypotheses are incorrect, except:

A Both hypotheses can possibly be true
B In some cases neither of the hypothesis can be true
C Exactly one hypothesis must be true
D Both hypotheses must be true

Question 5

Nutrient, a baby milk powder, is sold in tetra packs with an average weight of 1 kg. The standard deviation of the weight was calculated as 75 gm.

Few of the regular customers desire to check the precision of the advertised average weight of the milk powder packets and take a sample of 50 packets finding an average weight of 0.9 kg.

What is the set of hypotheses that should be used to test the accuracy of advertised weight?

A $H_0: \mu = 1$; $H_1: \mu \neq 1$
B $H_0: \mu = 1$; $H_1: \mu < 1$
C $H_0: x = 0.9$; $H_1: x \neq 0.9$
D $H_0: x = 1$; $H_1: x < 1$

Question 6

When can the chi-square goodness-of-fit test be used?

A To test for normality
B To test for probability
C To test for significance of sample statistics
D To test the difference between population means

Question 7

On which of the following assumptions ANOVA test is based on?

A Population is normally distributed
B Sample is randomly selected
C Population variance is equal to some common variance
D Options A, B and C

Answers to Self Examination Questions

Answer to SEQ 1

The correct option is D.

The null hypothesis is assumed to be true unless one find evidence to the contrary.

Answer to SEQ 2

The correct option is C.

In a hypothesis test, a Type I error occurs when the null hypothesis is rejected when it is in fact true; that is, H_0 is wrongly rejected.

Answer to SEQ 3

The correct option is A.

Usually, the significance level of a statistical hypothesis test is a fixed probability and is chosen to be 5%

Answer to SEQ 4

The correct option is C.

From null and alternative hypothesis, exactly one must be true.

Answer to SEQ 5

The correct option is A.

Answer to SEQ 6

The correct option is A.

The chi-square goodness of fit test can be used to test normality.

Answer to SEQ 7

The correct option is D.

ANOVA (analysis of variance) test is based on all the given assumptions.

LINEAR PROGRAMMING

8

Get Through Intro

Everyone wants a mansion as a house, a servant who anticipates our every request and a holiday that never ends. However, one matter puts a stop to all the daydreaming: our limited financial resources. Probably the only way you can have all this is if you have an “Aladdin’s genie” at your service!

Likewise a producer of goods always wants to fulfil the demands of the market as far as possible and make as much profit as he can. However, in reality resources are limited and this compels the producer to operate within the given constraints. These limited resources are known as limiting factors.

This Study Guide takes you through the process that allows one to earn the maximum possible profit even when there are limiting factors. It also explains the use of graphs and simultaneous equations to arrive at an optimal production plan.

Learning Outcomes

a) Formulate a Linear Programming problem.
b) Solve a linear programming problem by both graphical method and simultaneous equation method.
c) Obtain shadow/dual values.
d) Conduct sensitive analysis and explain slack variable.
e) Formulate a dual model from a paired model.
f) Determine shadow values.
g) Interpret shadow/dual values.
h) Apply the concept of linear programming in accounting and business situations.

1. Formulate a Linear Programming problem.

[Learning Outcome a]

A limiting factor (also known as principle budget factor) is any resource that limits and restricts the volume of production or it is the factor that is in short supply. A limiting factor restricts the indefinite expansion of the production or service activity. This in turn leads to restriction on earning and hence limits infinite profits. So far we have come across situations involving only one limiting factor. In situations where there is more than one limiting factor or scarce resources we have to resort to a method called "linear programming", in order to determine the optimal production plan.

1.1 The meaning of linear programming

Definition

Linear programming is a mathematical technique used to arrive at an optimal production plan and to achieve the optimal utilisation of resources like money, manpower, machines and so on, with an objective of either cost minimisation or profit maximisation.

The word programming is used because a program / routine is set up that can be followed to solve the problems of resource allocation. The major characteristics of the linear programming technique are:

1. The objectives (cost minimisation or profit maximisation) and the limitations or constraints (limiting factors) can be expressed as linear functions.
2. It provides scope to choose amongst the alternatives.
3. One or more of the factors of production must have restrictions in availability or use.
4. The variables in the objective function and the constraints are expressed in common units of measurement, (e.g. Tshs, $ or £) even if two products have different units of measurement.

Example

A dairy farm produces both butter and milk. These two products have different units of measurement i.e. kilograms and litres. Yet the objective function and constraints are expressed in terms of money so as to remove the difficulty of adding two distinct units of measurement (litres + kilograms in this case).

We have come across linear equations in the previous Study Guides of correlation and regression and time series analysis, where we expressed the costs as a linear equation involving variables.

In this Study Guide we will be dealing with equations involving two variables. Each variable represents the quantum of each of the products to be manufactured.

1.2 Steps involved in the formulation of a linear programming problem

1. Formulation of an objective function

Definition

Objective function is the objective of maximising profits or minimising cost, expressed in the form of a linear equation. The equation indicates the relationship between the output and the profit.

The above function, formulated in two variables, defines the purpose of the decision. It expresses the desired outcome in a mathematical form. An objective function is expressed as a combination of variables. The variables are the number of units of products to be produced.

Each variable is multiplied by either the profit per unit or the cost per unit, depending on whether the objective is to maximise profits or minimise costs.

2. Identification of constraints

In this step the possible limiting factors that restrict production are converted into mathematical equations similar to the objective function e.g. the limited labour hours available are 25,000, and product x consumes 5 (labour) hours per unit, whereas product y consumes 3 (labour) hours per unit. The constraint function will be formed as 5x + 3y $ 25,000. This constraint signifies that both products together should not consume more than 25,000 labour hours.

Let us try to formulate a linear programming problem.

Example

Green Tree Plc manufactures two types of manure, chemical and bio-chemic. Production is at full capacity and labour hours are limited to 2,500 for the year. The manure needs to be packed in a special packaging, which is produced by another company. This year they can supply Green Tree Plc with only 5,000 packets. The profit per packet of the chemical manure is Tshs5,000 and that of the bio-chemic manure is Tshs4,000. The labour hours required per packet are 2 hours for the chemical manure and 3 hours for the bio-chemic manure.

In the first stage, we will identify the objective function, which defines the purpose of solving the question. It can either be profit maximisation or cost minimisation.

In the above example the profit per unit for the products is given and hence the objective is to maximise profit.

Linear programming makes use of equations and / or inequalities involving two or more variables to solve the problem. We assign variables to the values that need to be ascertained. We need to ascertain the optimal production plan, meaning the number of units that need to be produced. We then assign variables to the number of units to be produced.

Let us assign the variable ‘x’ to the packets of chemical manure to be produced, and variable ’y’ to the packets of bio-chemic manure to be produced. The objective function can be expressed as follows:

Maximise (Profit per packet of chemical manure x x + profit per packet of bio-chemic manure x y)

The objective function will be expressed as: maximise 5x + 4y (Maximise Z = 5000x + 4000y)
Where Z is the profit function

The second stage is to identify and define the constraints.

The first limiting factor is the limited labour hours available. The per unit requirement of labour hours is 2 hours for the chemical and 3 hours for the bio-chemic manure. The hours available are 2,500. So the first constraint can be put as:

Hours required per packet of chemical manure x number of packets + hours required per packet of bio-chemic manure x number of packets should be “less than or equal to” the total 2,500 hours available.
2x + 3y $ 2,500

The next constraint is the amount of packaging material available. Since the product is incomplete without the proper packaging we can produce only an amount for which there is proper packaging available.
Therefore the constraint is x + y $ 5,000

The variables cannot be in the negative as the number of packets cannot be negative, so the other constraints can be identified as: x ;: 0 and y ;: 0

This is known as the non-negativity constraint which applies to all the problems.

We have formulated the linear programming problem as follows:

Maximise the objective function 5x + 4y (this means maximise the profits).(Maximise Z = 5000x +4000y)

Subject to constraints 2x + 3y $ 2,500 (this means that labour hours must be less than 2,500hrs and split between the 2 products).

x + y $ 5,000 (there is a maximum of 5,000 plastic covers (packets available for packing), which must be split between the 2 products).

x ;: 0 and y ;: 0 (all production should be positive)

Similarly one can formulate the linear programming problem for any situation involving constraints. These equations can be solved to ascertain the values of x and y which will give us the optimal production plan.

Let us consider another simple decision situation.

Example

Arusha Cattle Feed Ltd manufactures two varieties of cattle feed – fodder plus and fodder premium. The details of the processing time in the mixing and blending departments are given below, as is the profit per kilogram. Formulate a linear programming problem to arrive at the optimal product mix.

	Mixing department	Blending department	Profit per unit (Tshs'000)
(Fodder plus	5 hours	4 hours	4
Fodder premium	6 hours	1.5 hours	2
Total time available	30 hours	24 hours	

Answer

Mixing department constraint

5 hours are required to produce a unit of fodder plus, therefore the hours required to produce x units will be the number of units x hours required = 5x

6 hours are required to produce a unit of fodder premium, therefore the hours required to produce y units will be number of units x hours required = 6y

The total hours available are 30
The constraint will be 5x + 6y $ 30 hours

Blending department constraint

4 hours are required to produce a unit of fodder plus, therefore the hours required to produce x units will be 4x

1.5 hours are required to produce a unit of fodder premium, therefore the hours required to produce y units will be 1.5y

The total hours available are 24
The constraint will be 4x + 1.5y $ 24 hours
The linear programming problem will be:

Maximise the objective function	4x + 2y
Subject to constraints	5x + 6y $ 30 hours
	4x + 1.5y $ 24 hours
The non-negativity constraint	x, y ;: 0

Test Yourself 1

In which of the following case, the linear programming problem applies?

A A situation where there is more than one limiting factor
B A situation where the limiting factors do not apply.
C A situation where only a single product is manufactured.
D None of the above

Test Yourself 2

In a given linear programming problem the constraint is that twice the production of the product denoted by x cannot exceed three times the production of the product denoted by y. The constraint will be expressed as:

A 2x > 3y
B 2x $ 3y
C 3x ;: 2y
D 3x < 2y

2. Solve a linear programming problem both by graphical method and simultaneous equation method.

[Learning Outcome b]

Linear programming problems can be solved using graphs when the equations / inequalities are expressed in two variables. Any equation / inequality containing more than two variables cannot be plotted as the graph consists of only two axis – X and Y.

2.1 The steps of formulating a linear programming model in multiple scarce resource situations are as follows:

1. Determining the objective variable and objective function

An objective variable represents the quantity given ((the notation x_1, x_2, x_3, etc.) of an alternative item or course of action that is required to attain the optimum solution.

An objective function is the quantification of an objective, usually expressed as maximising profit or minimising cost. It is a structured description of what one wants to do.

An objective function expressed in a structured form but in English will be:

Maximise contribution margin or Minimise costs

To solve a linear accounting problem, the objective function needs to be expressed in terms of algebraic symbols, such as

Maximise $Z = c_1x_2 + c_2x_2 + \ldots\ldots\ldots\ldots + c_nx_n$; where, c_1, c_2, etc. are constants representing per unit contribution to the objective function.

Similarly, an objective function may be constructed for minimising costs also when c_1, c_2, etc. would signify costs per unit.

Example

East African TV Company produces two types of TV sets, 'Romantic' and 'Elegant'. The company has a capacity to produce 60 sets per day of the 'Romantic' and 50 sets per day the 'Elegant'. The labour requirement for the 'Romantic' set is 1 man-hour, whereas 'Elegant' requires full 2 man-hours of labour.

Presently, there is a maximum of 120 man-hours of labour per day that can be assigned to the production of the two types of sets. The contribution margins are Tshs20,000 and Tshs30,000 for each 'Romantic' and 'Elegant' set respectively. Given these circumstances, the company wants to know what the daily production of its products should be.

Here, the objective function may be expressed mathematically as follows:

Maximise $20{,}000X_1 + 30{,}000X_2$

When contribution margin is measured in shillings which are Tshs20,000 and Tshs30,000 respectively for Romantic and Elegant, and X_1 and X_2 represent the quantity of Romantic and Elegant(tv sets) to be produced in a day.

---

2. Constructing constraint statements

The multiple scarce resources, i.e. limiting factors, need to be expressed as linear inequalities or equations in terms of the above defined decision variables.

The constraint statement can be expressed using algebraic symbols as:

$a_{11}x_1+a_{12}x_2+ \ldots\ldots\ldots\ldots + a_{1n}x_n \$ \text{ or } ;: b_1$

$a_{21}x_1+a_{22}x_2+ \ldots\ldots\ldots\ldots + a_{2n}x_n \$ \text{ or } ;: b_2$

$a_{m1}x_1+a_{m2}x_2+ \ldots\ldots\ldots\ldots + a_{mn}x_n \$ \text{ or } ;: b_m$

$x_1, x_2 \ldots\ldots\ldots\ldots x_n ;: 0.$

where b_1, b_2,..............,b_m etc. are indicative of constraint requirements or availability.

Example

Continuing the example of East African TV Company

The constraints may be expressed in English as follows:

Capacity constraint
Romantic's production less-than-or-equal-to capacity for Romantic measured in units)
Elegant's production less-than-or-equal-to capacity for Elegant measured in units)

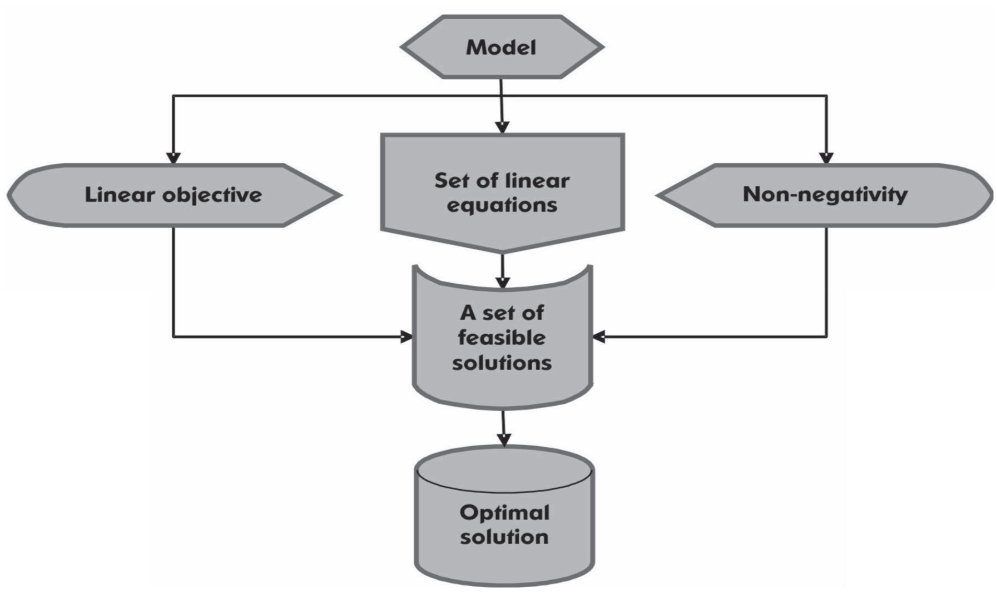
Labour Hour constraint
Labour used less-than-or-equal-to labour availability measured in man-hours)

The can be expressed as follows:
$X_1 \$ 60$ (Romantic's capacity)
$X_2 \$ 50$ (Elegant's capacity)
$X_1+2X_2 \$ 120$ (Labour in man-hours)

It is very common to denote two variables as either X_1, X_2 or X, Y or x, y

Where, X_1, X_2 ;: 0 this is the non-negativity assumption)

Diagram 2: Ingredients of LP model



3. Solve a multiple scarce resource problem graphically

A multiple scarce resource problem can be solved graphically only if the number of decision variables is two or less. On the other hand, the graphical method of solving a linear programming problem can deal with any number of constraints but as each constraint is represented by a line on the graph, a large number of lines may make the graph difficult to read.

Despite the above characteristic features, graphical methods are the simplest methods of solving a linear programming problem and are therefore used extensively.

The method of solving a linear programming problem graphically is as follows:

(a) Formulate the appropriate linear programming problem.

(b) Treat the inequalities as equalities and draw a line for each equation on a piece of graph paper. To draw the lines representing the equations, select the two points that intersect the axes and connect them with a straight line.

(c) Note that a constraint of the 'less than' type will have a feasible region on its left or below the line. Conversely, in the case of a constraint of the 'more than' type, the feasible region will be on the right hand side of the line representing the constraint. The feasible region also includes the line concerned, if it is an equation or an inequality of the 'less than or equal to' or the 'more than or equal to' type.

(d) Identify the solution space or feasible region that satisfies all the constraints simultaneously. The feasible region must be bounded by the lines drawn and the axes. If the region is not bounded, it implies the feasible region is unbounded; the linear programming problem has no solution when the there is no common feasible region available for all the constraints given in a linear programming problem)

(e) Evaluate the objective function at each of the corner points of the feasible region and accordingly, identify the optimal value of the objective function. It should be noted that the optimal solution lies at any of the corner points of the feasible region).

Example

The problem of East African TV Company is solved graphically in the following steps.

Step 1: Formulate the Linear Programming Problem as follows:

Objective function: Maximise $20{,}000X_1 + 30{,}000X_2$

Subject to Constraints
$X_1 \$ 60$ (Romantic's capacity)
$X_2 \$ 50$ (Elegant's capacity)
$X_1 + 2X_2 \$ 120$ (Labour in man-hours)

Where, X_1, X_2 ;: 0

Step 2: Remember, straight lines will have to be drawn for each constraint. And the decision variable of the objective function is shown on each of the x-axis and y-axis. We will show the variable 'X_1' on x-axis and 'X_2' on y-axis

For drawing the straight line, we must convert the inequalities into equalities as follows:

$X_1 = 60$
$X_2 = 50$ and
$X_1 + 2X_2 = 120$

Hence there will be in all three equation lines plotted on the graph.

The first line will be $X_1 = 60$ will go parallel to y-axis and the second line $X_2 = 50$ will be parallel to the X-axis. For the third line we will have to select the points and then join them to get the line for $X_1 + 2X_2 = 120$. This is shown below:

So as to obtain the points of intersection with the axes, put $X_1 = 0$ in the equation $X_1 + 2X_2 = 120$
i.e. $0 + 2X_2 = 120$

$X_2 = 60$

Again, put $X_2 = 0$ in the equation $X_1 + 2X_2 = 120$
i.e. $X_1 + 2 \times (0) = 120$

$X_1 = 120$

Alternatively, the equation $X_1 + 2X_2 = 120$ may be expressed into intercept form to be presented as

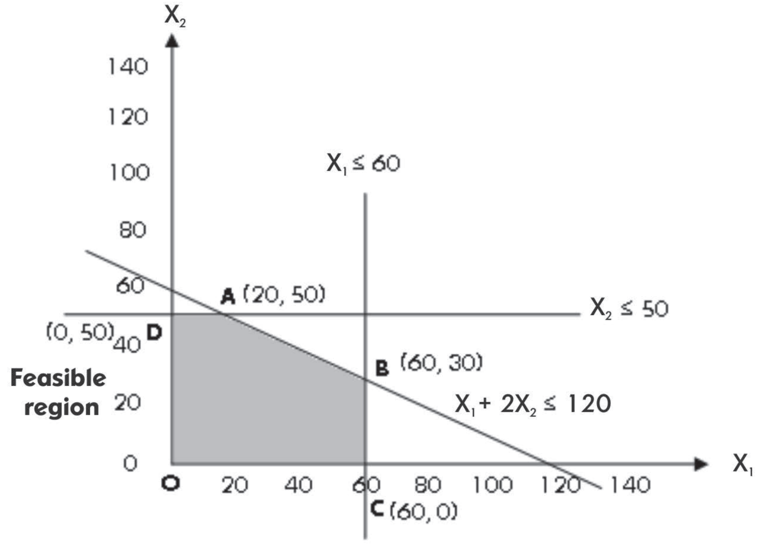
$$\frac{X_1}{120} + \frac{X_2}{60} = 1$$

by dividing both the sides of the equation by the constant term i.e. (120).

Accordingly, we can directly infer that the straight line intersects the horizontal axis at (120,0) and the vertical axis at (0,60) points.

Continued on the next page

Step 3: Graphical presentation of Linear Programming problem



Step 4: The shaded portion shown in the graph reflects the feasible solution area i.e. ODABC

Find out the co-ordinates of all corner points of the feasible region

Co-ordinates of point A

The $X_2 = 50$ line intersected the $X_1 + 2X_2 = 120$ line. Therefore, by putting the value of X_2 into the equation $X_1 + 2X_2 = 120$, we get $X_1 = 20$
Therefore, the co-ordinates of point A are (20,50)

Co-ordinates of point B

The $X_1 \$ 60$ line intersected the $X_1 + 2X_2 \$ 120$ line. Therefore, by putting the value of X_1 into the equation $X_1 + 2X_2 \$ 120$, we get $X_2 = 30$
Therefore, the co-ordinates of point B are (60,30)

Co-ordinates of point C

Find out the points of intersection of the equation $X_1 \$ 60$
First this inequality is converted into the equation $X_1 = 60$
The equation line is parallel to X_2 i.e. this line is at a 60 units distance from the X_2 axis)
Therefore, the co-ordinates of point C are (60,0)

Co-ordinates of point D

Find out the points of intersection of the equation $X_2 \$ 50$
First this inequality is converted into the equation $X_2 = 50$
The equation line is parallel to X_1
Therefore, the co-ordinates of point D are (0,50)

Step 5: Find out the optimum profit of East African TV Company by putting the value of the coordinates of the corner points of the feasible region into the objective function $Z = 20X_1 + 30X_2$ (assuming Tshs'000)

Corner Point	Coordinates of corner point	Objective function $Z = 20X_1 + 30X_2$ (Tshs'000)	Value (Tshs'000)
O	(0, 0)	Z (O) = 20 x(0) + 30 x (0)	0
D	(0, 50)	Z (D) = 20 x(0) + 30 x (50)	1,500
C	(60, 0)	Z(C) = 20 x (60) + 30 x (0)	1,200
A	(20, 50)	Z (A) = 20 x (20) + 30 x (50)	1,900
B	(60, 30)	Z (B) = 20 x (60) + 30 x (30)	2,100

Out of the five corner points, the fifth point i.e. for point (B), gives the maximum profit of Tshs2,100,000 indicating the optimum profit for the business. The optimum profit will be achieved when 60 units per day of the Romantic model are produced and 30 units per day of the elegant model are produced.

2.2 Iso-contribution line

Iso-contribution line is actually a short-cut method to find out the optimal solution. Iso means similar and iso-contribution line means a line that is similar (but not exactly the same) to a contribution line, i.e. a line that has the same slope as a contribution line would have.

Example

In the above example of East African TV Company, the objective function (i.e. the contribution function) is $20X_1 + 30X_2$. To find out the iso-contribution line, multiply the coefficient of the two variables together.

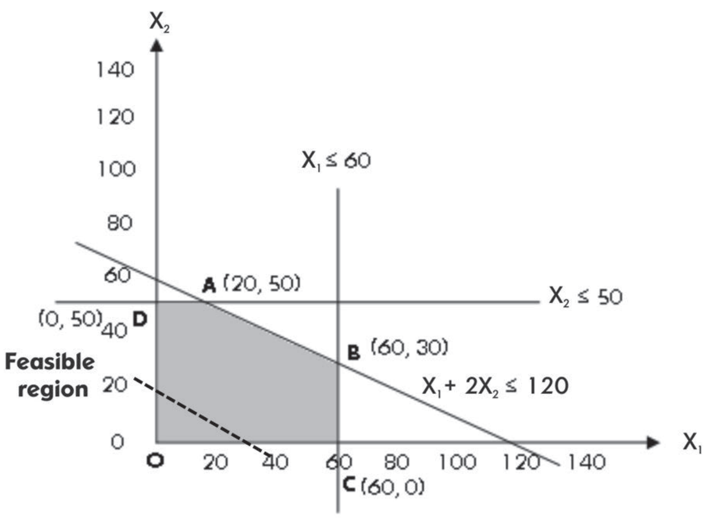
In our contribution function, i.e. $20X_1 + 30X_2$, X_1 and X_2 are two variables and the coefficient of X_1 is 20 whereas the coefficient of X_2 is 30.

Simply multiply these coefficients (20x30) to get the number that is to be used to convert the function into an equation. This will be the equation of the iso-contribution line.

Therefore: $20X_1 + 30X_2 = 600$

We now have the equation of the iso-contribution line. We can use this to find the coordinates of X_1 and X_2 and draw the iso-contribution line.

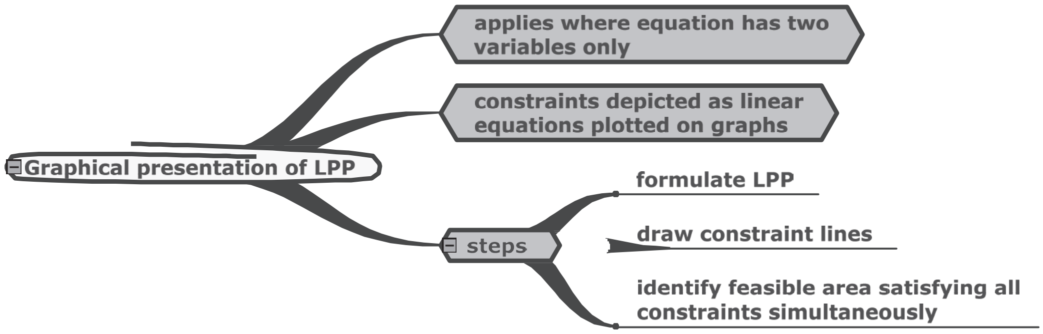
By solving the above equation:
$X_1 = 30$ when $X_2 = 0$
$X_2 = 20$ when $X_1 = 0$



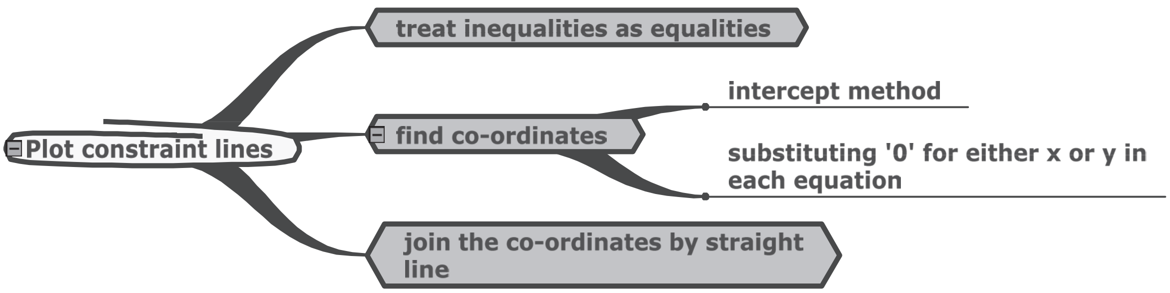
The optimal point can be easily determined using the iso-contribution line. You need to place a ruler on this line and move slowly up parallel to the line. While moving up, the ruler will touch the points in the feasible region ODABC one by one. The last point that is touched before leaving the feasible region is called the optimal point.

It is very important to note that you need to move the iso-contribution line outwards for profit maximisation and inwards for cost minimisation.

SUMMARY



SUMMARY



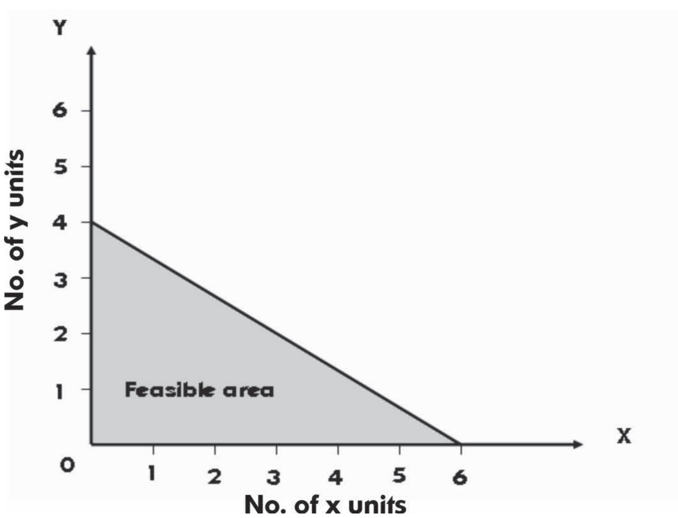
Test Yourself 3

What is the feasible area in a linear programming problem?

A The area where the best possible profits can be achieved within the given constraints
B The area where the minimum possible profits can be achieved within the given constraints
C The area where the maximum possible losses can be avoided
D All of the above.

Test Yourself 4

The graph given below is drawn for the constraint



A 2x + 3y $ 12
B 5x + 4y $ 20
C y $ 20
D x $ 10

Test Yourself 5

Timo Ltd makes only two products. The production constraints representing two machines and their maximum hours available are as follows:

Machine 1: 2X + 3Y $ 18
Machine 2: 2X + Y $ 10

Where X = Number of units of the first product
Y = Number of units of the second product

If the objective function is to maximise total contribution margin of Tshs4,000X + Tshs3,000Y, the most possible contribution margin is:

A Tshs24,000
B Tshs20,000
C Tshs18,000
D Tshs27,000

Simultaneous equations can be used to solve a linear programming problem with two variables and a minimum of two constraints.

Consider the following situation:

A manufacturer produces two products 'Instant' and 'Quick'. These two products require different times on two machines M1 and M2. The time required for Instant on M1 is 3 hours and on M2 is 1 hour. The time required for Quick on M1 is 1 hour and on M2 is 2 hours. The profit per unit of Instant is $4 and per unit of Quick is $5. The restricted machine hours for M1 are 9 hours and for M2 are 8 hours per day. The manufacturer wants to know how many units should be produced each day to maximise profits. Formulate a linear programming problem and arrive at an optimal production plan, using simultaneous equations.

Let x be the number of units of Instant to be produced and y be the number of units of Quick to be produced.

Example

The profit is given as Tshs4,000 for Instant and Tshs5,0000 for Quick. The objective function will be to maximise profit 4,000x + 5,000y.

Here we have assume X_1 as x and X_2 as y

Z = 4x + 5y (Amounts in tshs'000)

The hours are restricted to 9 hours on machine M1 and 8 hours on machine M2. The constraints can be written as:
For machine M1: 3x + 1y $ 9 hours
For machine M2: 1x + 2y $ 8 hours

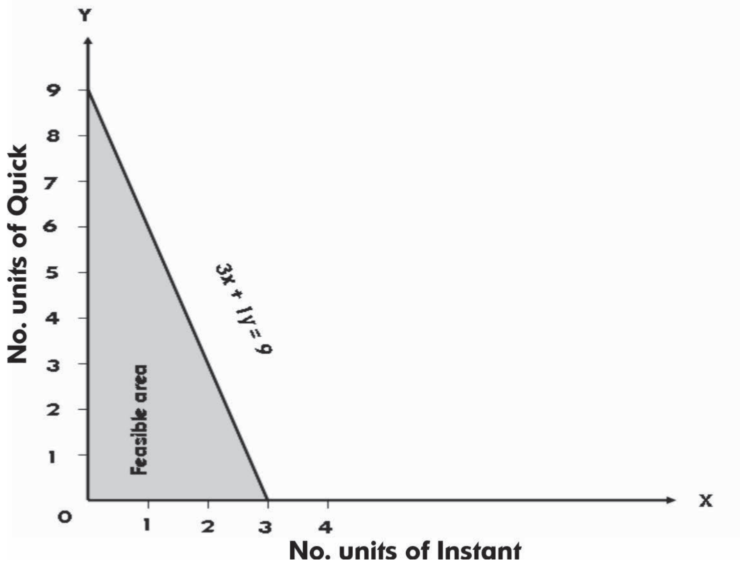
Let us draw the constraints on the graph
To plot the co-ordinates of the constraint 3x + 1y $ 9
3x + 1y = 9
Substituting y = 0 we have
3x = 9
x = 3
Co-ordinates - (3, 0)

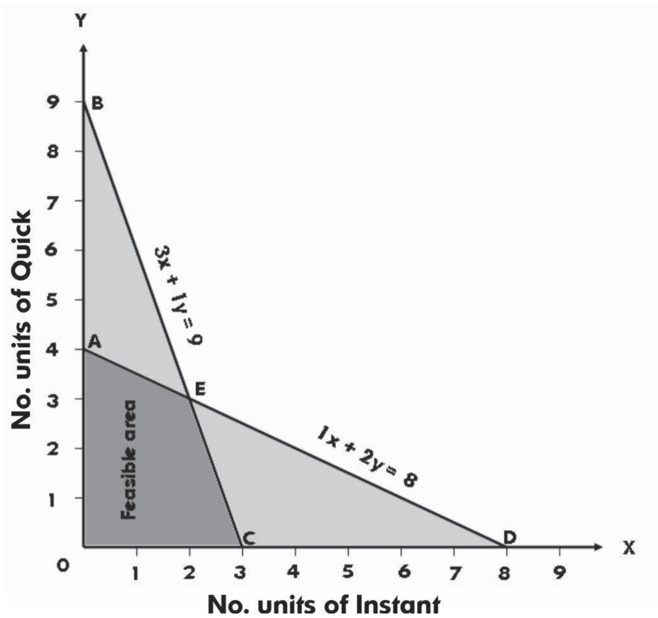
Substituting x = 0 we have
1y = 9
y = 9
Co-ordinates (0, 9)

For the constraint 1x + 2y $ 8
1x + 2y = 8
Substituting y = 0 we have 1x = 8, x = 8 Co-ordinates - (8, 0)
Substituting x = 0 we have 2y = 8, y = 4 Co-ordinates – (0, 4)

Continued on the next page

The graph can be plotted as below:





The co-ordinates for the point where the constraint lines intersect can be calculated as follows:
The equations for the lines are:
3x + 1y = 9 ……… (1)
1x + 2y = 8 ……… (2)

Multiplying equation (2) by 3 we get:
3x + 1y = 9 ……… (3)
3x + 6y = 24 …… (4)

Subtracting equation (3) from (4)
5y = 15
y = 3

Substituting the value of y in equation (1) we have
3x + 1(3) = 9, x = 2
The co-ordinates are (2, 3)

The other co-ordinates from the graph can be plotted as (3,0) , (0,9) , (8,0) and (0,4).
However, the co-ordinates for the feasible region are OABC – (OAEC)

O (0, 0)	A (0, 4)
E (2, 3)	C (3, 0)

Continued on the next page

When we substitute the values of x and y as calculated, for the feasible area in the objective function we have the following solutions:

4,000x + 5,000y
At O - (0, 0) = 4,000(0) + 5,000(0) = 0
At A - (0, 4) = 4,000(0) + 5,000(4) = 20,000
At E - (2, 3) = 4,000(2) + 5,000(3) = 23,000
At C - (3, 0) = 4,000(3) + 5,000(0) = 12,000

The maximum profit is at point E Tshs23,000 at the co-ordinate E (2, 3). Therefore, the manufacturer should manufacture 2 units of Instant and 3 units of Quick to attain the maximum profit.

The simultaneous equations can be useful to ascertain the values of the co-ordinates for the points where two constraint lines intersect each other. It generally happens that the optimal solution lies at the point of intersection of the two constraint lines. Simultaneous equations are a much quicker and easier way to find the solution than the graphical approach. However, simultaneous equations do not show the values of all the corner points.

When the linear programming problem contains only two constraints the simultaneous equation method can provide the answer without plotting the graph.

Example

In the above case with the given data we have the equations:

Maximise 4,000x + 5,000y
Subject to constraints:
For machine M1: 3x + 1y $ 9 hours

For machine M2: 1x + 2y $ 8 hours

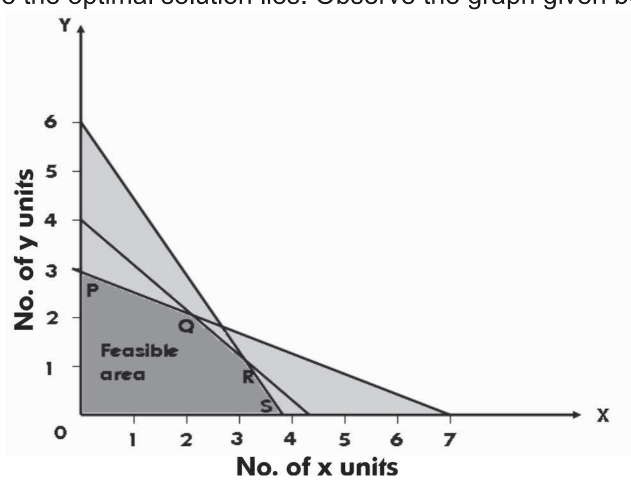
Assuming these inequalities to be equalities we have

3x + 1y = 9 …………… (1)
1x + 2y = 8 …………… (2)

Solving these two equations in the same way as above, we can ascertain the co-ordinates at the point of intersection of these two constraint lines - (2, 3).

Since the optimal solution generally lies on the point of intersection, we may directly ascertain the co-ordinates at the point of intersection, to arrive at the optimal solution.

However, where there are more than two constraints you will need to conduct several trial and errors to arrive at the optimal product mix. This is because as the number of constraint lines increases, the number of possible intersections will also increase. In this case, without drawing the graph we will not be in a position to ascertain the point of intersection where the optimal solution lies. Observe the graph given below.



Continued on the next page

In the above graph there is more than one point of intersection where the feasible solution may lie. As a result, on plotting the graph we may directly solve the equations for those constraint lines whose points of intersection lie in the feasible area. In the above case the equations need to be solved for points Q and R so as to arrive at a feasible solution.

Test Yourself 6

MO Plc manufactures two products which have to pass through three departments. The time consumed by the products in each of the processes is as follows:

	Product P (hr / unit)	Product Q (hr / unit)	Available capacity (hours)
Process I	8	8	560
Process II	2	4	200
Process III	5	2	320

The profit per unit of P is Tshs4,000 and that of Q is Tshs3,000. Formulate a linear programming problem and arrive at the best product mix. Make use of the simultaneous equations approach.

3. Obtain shadow/dual values.
   Formulate a dual model from a paired model.
   Determine shadow values.
   Interpret shadow/dual values.
   Apply the concept of linear programming in accounting and business situations.
   [Learning Outcomes c, e, f, g and h]

3.1 Primal and dual problems

Every linear programming problem has an equal but opposite formulation.

According to the duality theorem, "for every maximisation or, minimisation) problem in linear programming, there is a unique similar problem of minimisation or maximisation) involving the same data which describes the original problem".

The same solution can be obtained from either formulation. The original problem is termed the primal problem and the equal but opposite formulation is termed the dual problem.

Example

Food P contains 20 units of vitamin B and 40 units of vitamin C per gram. Food Q contains 30 units each of vitamins B and C. The daily minimum human requirements of vitamins B and C are 900 milligrams and 1,200 milligrams respectively. The primal problem is to quantify how many grams of each type of food should be consumed so as to minimise the cost if food P costs 60 cents per gram and food Q costs 80 cents per gram.

The buyer's primal problem may be formulated as follows:
Minimise $Z = 60x_1 + 80x_2$

Subject to the constraints,
$20x_1 + 30x_2$;: 0.9 (0.9 grams are equivalent to 900 milligrams)
$40x_1 + 30x_2$;: 1.2 (1.2 grams are equivalent to 1200 milligrams)
x_1, x_2 ;: 0

Moreover, the seller wishes to fix the maximum per unit selling price of the two vitamins B and C in such a way that the prices of foods P and Q do not exceed their existing price per unit in the market.

The seller may designate variables y_1 and y_2 to represent per unit price and assign these variables to vitamins B and C respectively, while making his selling plan.

Continued on the next page

The seller's dual problem note that the coefficients of the dual variables have been obtained by transposing the coefficient matrix of the primal may be formulated as:

Maximise z' = $0.9y_1$+ $1.2y_2$
Subject to the constraints,
$20y_1 + 40y_2$ $ 60
$30y_1 + 30y_2$ $ 80
y_1, y_2 ;: 0

3.2 Conceptualising shadow price

Definition

A shadow price is an increase in value which would be created by having available one additional unit of a limiting resource at its original cost. This represents the opportunity cost of not being able to use the one extra unit.

The shadow price is the change in the objective value of the optimal solution of an optimisation problem determined by relaxing the scarce resources by one unit. Each constraint in an optimisation problem has a shadow price or dual variable.

Example

If a business has a constraint that limits the labour availability to 200 hours per week, the shadow price will indicate how much the management of the business should be willing to pay for an additional hour of labour.

If the shadow price is Tshs30,000 for the labour constraint, one should pay no more than Tshs30,000 an hour for additional labour. Labour costs of less than Tshs30,000 per hour will increase the contribution margin objective value); labour costs of more than Tshs30,000 per hour will decrease the contribution margin objective value). Labour costs of exactly Tshs30,000 will cause the contribution margin to remain the same.

A scarce resource only has a shadow price when it is binding. Where resources are not fully utilised i.e., when the resources are in surplus) the shadow price is zero. This supports the logic that there can be no benefit in increasing the quantum of a resource of which there is already a surplus.

3.3 Calculation of shadow price

Shadow prices are calculated automatically under the Simplex method a systematic method of solving linear programming problems through iterations) of solving linear programming problems. However, whenever shadow prices of problems need to be calculated without resorting to the Simplex method, one can do this by solving the dual inequalities of a primal problem.

Example

Continuing the above of example

The shadow price (dual price) can be obtained by solving the constraint inequalities considering that they are equations for the values of the variables of the dual problem.

The shadow prices are found as follows:
If y_1 = shadow price for vitamin B,
And y_2 = shadow price for vitamin C

The dual constraints become
$20\ y_1 + 40\ y_2 = 60$ ……….(equation 1)
$30\ y_1 + 30\ y_2 = 80$ ……….(equation 2)

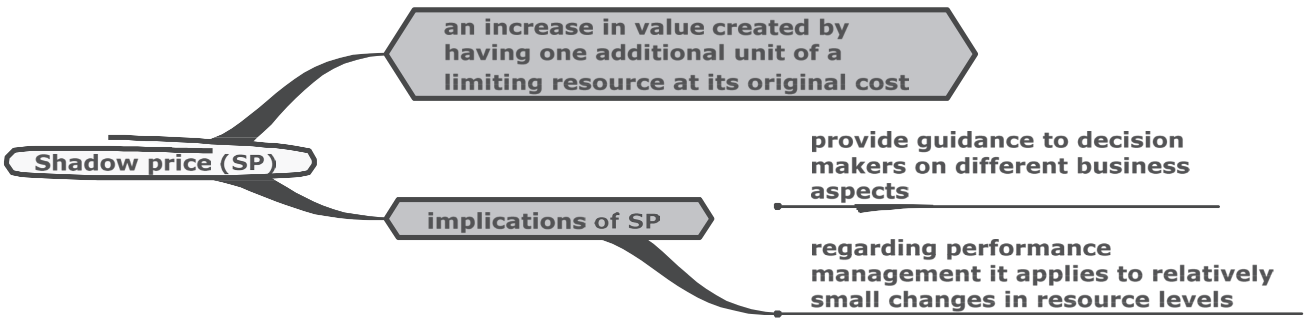
Solving the equations, $y_1 = 2.33$
$y_2 = 0.33$

Therefore the shadow prices are 2.33 cents and 0.33 cents for vitamins B and C respectively and the contribution to the seller comes to 0.9 x 2.33 + 1.2 x 0.33 = 2.493 cents.

3.4 Implications of shadow pricing on decision-making and performance management

The shadow price of a resource is the opportunity cost of that resource. Shadow price can provide guidance to decision-makers on different business aspects such as the value to the business of relieving existing constraints. However, shadow price normally applies to relatively small changes in resource levels. If the change in the amount of the resource is of a significant magnitude then the problem should be resolved using the revised resource levels when the new shadow price emerges. The shadow price can provide decision-makers a powerful insight into the problems.

SUMMARY



Test Yourself 7

A company makes two kinds of leather belts. Belt A is a high quality belt, and belt B is of lower quality. The respective profits are Tshs40,000 and Tshs30,000 per belt. Each belt of type A requires twice as much time to make as a belt of type B, and if all belts were of type B, the company could make 1,000 per day. The supply of leather is sufficient for only 800 belts per day both A and B combined). Belt A requires a fancy buckle, and only 400 per day are available. There are only 700 buckles a day available for belt B.

Required:

Formulate the linear programming problem (primal) and also formulate its dual problem.

Test Yourself 8

Using the following information, calculate the shadow price for assembly time.

Maximise z = 1.5x + 1.8y (Amounts in Tshs'000)
Subject to constraints,
15x + 24y $ 600 (cutting time)
30x + 24y $ 720 (assembling time)

Application of linear programming

As seen in the previous examples, linear programming is widely applied in the areas of decision making where limiting factors exists.

The following is the list of examples where linear programming technique is applied:

1. It is used in transportation problems. In fact, transportation and assignment problems are considered as structured linear programming. In case, there are various supply at different locations and various destinations where goods need to be supplied, this technique helps to minimise cost of transportation
2. Assignment, a part of transportation, is also based on linear programming technique. Where there are multiple resources and multiple tasks to be performed, this model helps in optimal allocation of resources.
3. LP technique mainly aims for optimisation. In case of cost minimisation of revenue maximisation, it helps the decision maker make the optimal decision by using sensitivity analysis.
4. It is also very useful in diet problems. What amount of vitamin and calcium / mineral should be consumed to get maximum health benefit can be calculated using LP technique.
5. Nowadays, LP technique is very much useful for investment companies as they have also started considering optimisation.

4. Conduct sensitive analysis and explain slack variable.
[Learning Outcome d]

4.1 Sensitivity Analysis

1. Meaning

While appraising the profitability of any project over a period, certain assumptions are made about the parameters affecting profitability. However, these assumptions may not turn out to be true in reality. As a result, a project may not achieve the anticipated results. Sensitivity analysis studies the impacts of these parameters on the effectiveness of the project and, moreover, the extent of the impacts.

Sensitivity analysis is a "what–if" technique that measures how the expected values in a decision model will be affected by changes in the critical data inputs. In the context of cost-volume-profit analysis, sensitivity analysis answers questions such as "what will my net income be if the unit variable costs or the sale prices change by some amount from the original prediction?"

Sensitivity analysis explores the various risks to the project and their impact on the effectiveness of the project.

If a small change in a factor results in a relatively big change in the outcome, the project is said to be sensitive to that factor.

Example

If an increase in the labour rate affects the profitability of the project to a great extent, but a change in material cost per unit does not affect profitability substantially, then one may conclude that the project is very sensitive to the labour rate but not sensitive to material cost.

Sensitivity analysis involves examining a project in the light of possible changes in the parameters of the project. From this exercise, the decision-makers may find out about the margin they should leave for judgemental error. This may allow them to decide whether they are prepared to bear the associated risks or not.

Sensitive analysis provides decision-makers with information about the responsiveness of the project to the parameters. Once the crucial factors have been determined, the decision-makers can concentrate only on those crucial factors, and thereby ensure the sustainability of the project.

Example

Venture Ltd has identified that its new project is very sensitive to labour cost. Therefore, while deciding where to locate its plant, it will make sure that the location is a place where cheap labour is available.

Sensitivity analysis influences management to make contingency plans, if the project is highly susceptible to change for a factor critical to it.

Example

Continuing the above example of Venture Ltd, as the project is sensitive to labour cost, management may have a contingency plan for outsourcing, in case an adverse situation arises.

Another major drawback of sensitivity analysis is that, even though this analysis is based on probabilistic events, the absence of a formal assignment of probabilities to the variations of the parameters limits the effectiveness of the technique.

Example

Even if it is determined that the project is sensitive to labour cost, it is impossible to determine when labour cost will increase or decrease or to what extent it will increase or decrease.

---

2. Application of sensitivity analysis

Example

Symsoft Plc has developed a software package. The demand for the package is estimated at 8,000 per year at a price of Tshs50,000. The company is also developing an advanced version of the software package and it is estimated that, when this new software package becomes available on the market next year, the company will stop selling the old version.

The estimated profit is as follows:

	Tshs'000	Tshs'000
Selling price		50
Less: Variable Costs		
Direct material	10	
Direct labour	30	
Overheads	05	
Total variable costs		45
Profit		5
Cash flow from the project is – Profit (Tshs5,000 x 8,000 units)		40,000

Before clearing this project, the project management committee wants to test the sensitivity of the project to the following changes in the parameters:

(a) A decrease in selling price of 5% causing an increase in demand of 2%
(b) An increase in labour cost of 7%
(c) A decrease in material cost of 2%
(d) An increase in selling price of Tshs2,000 causing a decrease in demand of 3%

Answer

(All amounts are in Tshs'000)

(a) A decrease in selling price of 5% causing an increase in demand of 2%

If the selling price goes down by 5%, the new price would be Tshs(50 - 2.5) = Tshs47.5.
If the volume increased by 2% (i.e. 160 units) the number of units sold would be 8,160.

The resultant profit would be:

	Tshs
Revenue (Tshs47.5 x 8,160)	387,600
Costs (Tshs45 x 8,160)	367,200
Profit	20,400

(b) An increase in labour cost of 7%

If the labour cost increases by 7%, the total cost would be
= (Tshs10 + 107% x Tshs30 + Tshs5)
= Tshs47.1

The resultant profit would be:

	Tshs
Revenue (Tshs50 x 8,000)	400,000
Cost (Tshs47.1 x 8,000)	376,800
Profit	23,200

(c) A decrease in material cost of 2%

If the material cost decreases by 2%, the total cost would be
= (Tshs45 - 2% x Tshs10)
= Tshs44.8

Continued on the next page

The resultant profit would be:

	Tshs
Revenue (Tshs50 x 8,000)	400,000
Cost (Tshs44.8 x 8,000)	358,400
Profit	41,600

(d) An increase in selling price of Tshs2 causing a decrease in demand of 3%

If the selling price increases by Tshs2, it would be
= Tshs50 + 2
= Tshs52

If the demand was to decrease by 3%, the number of units sold would be 7,760.

The resultant profit would be:

	Tshs
Revenue (Tshs52 x 7,760)	403,520
Cost (Tshs45 x 7,760)	349,200
Profit	54,320

4.2 Conceptualising slack variable

Linear programming problems must be converted into augmented form before being solved by the simplex algorithm. Slack variables are introduced to the left hand side of 'less than' type constraints so as to convert the inequality into an equation. Slack variables indicate that any constraint may remain unused at the point of optimality. If any resource or limiting factor is totally exhausted while attaining optimal solution, then there is no slack i.e., value of the slack variable is zero).

4.3 Calculate slack and explain the implications of the existence of slack for decision-making and performance management

The value of slack is derived automatically when the optimal solution to a linear programming problem is attained by a simplex algorithm. However, the value of slack can also be calculated when the solution to the problem is derived by other methods, namely the graphical or simultaneous equation method.

After identifying the optimal solution, the optimal values of the decision variables are substituted to the left hand side of the constraints so as to obtain important information about whether a resource could be fully utilised or some part of the resource remained unutilised. Also, which of the constraints are of critical importance will be revealed.

Once this step is taken, two kinds of outcome are possible. First, the total value on the left hand side of the constraint may be equal to that on the right hand side indicating that the resource is fully utilised in the process of optimising contribution i.e., there is no slack. Second, the value on the left hand side of the constraint may be less than that on the right hand side signifying that the resource is not fully utilised i.e., there is slack.

The analysis highlights that the constraints that are fully utilised are critically important and obtaining more of these resources would enable an organisation to improve upon the optimal solution.

Example

A firm makes two types of quality furniture chairs and tables. The contribution for each product as calculated by the accounting department is Tshs20,000 per chair and Tshs30,000 per table. Both products are processed on three machines M_1, M_2, M_3. The time required in hours by each product and total times available in hours per week on each machine are as follows:

Machine	Chair	Table	Available time
M_1	3	3	36
M_2	5	2	50
M_3	2	6	60

The management of the firm wanted to schedule its production in order to maximise contribution. Accordingly, in an attempt to have an appropriate schedule of production, the management formulated the problem as follows:

Maximise total profit

$Z = 20,000X_1 + 30,000X_2$

Subject to the constraints: $3X_1 + 3X_2 \$ 36$

$5X_1 + 2X_2 \$ 50$
$2X_1 + 6X_2 \$ 60$
X_1 ;: 0, X_2 ;: 0

Where, X_1 = Number of units of chairs,
X_2 = Number of units of tables

On solving the problem graphically the optimum contribution was found to be $330. The optimum contribution was reached when 3 chairs and 9 tables were produced.

By substituting the objective variable in the constraints with the optimal values (i.e. 3 chairs and 9 tables), the following situations arise:

For machine M_1,

$3X_1 + 3X_2 \$ 36$
Or, $(3 \times 3) + (3 \times 9) \$ 36$
Or, $9 + 27 = 36$

This signifies that the full capacity of machine M_1 was utilised in maximising contribution. For this resource, i.e. time of machine M_1, there is no slack.

For machine M_2,

$5X_1 + 2X_2 \$ 50$
Or, $(5 \times 3) + (2 \times 9) \$ 50$
Or, $15 + 18 \$ 50$
Or, $33 \$ 50$

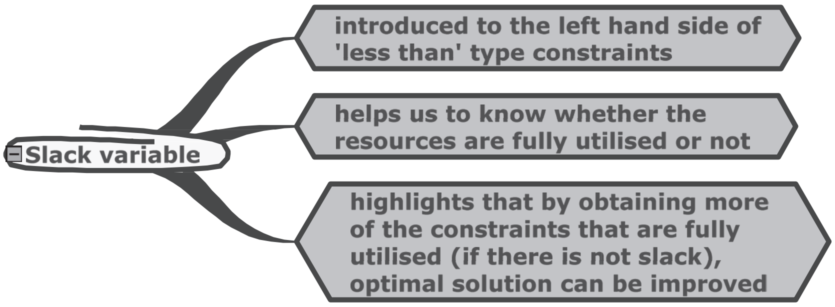
Here, the left hand side is less than the right hand side by 17, indicating that there are 17 units of slack or unused machine time.

For machine M_3,

$2X_1 + 6X_2 \$ 60$
Or, $(2 \times 3) + (6 \times 9) \$ 60$
Or, $6 + 54 = 60$

This signifies that the full capacity of machine M_3 was utilised in maximising contribution. For this resource, i.e. time of machine M_3, there is no slack.

SUMMARY



Test Yourself 9

Solve the problem of Tiny Ltd (discussed in Test Yourself 5) using the simultaneous equation method.

Test Yourself 10

Continuing Test Yourself 9, calculate slack for Tiny Ltd.

Answers to Test Yourself

Answer to TY 1

The correct option is A.

Linear programming is used when there is more than one limiting factor which can be expressed as linear equations. It won't apply if limiting factors do not apply. Single product manufacturing is also not suited to linear programming, as it involves only one variable.

Answer to TY 2

The correct option is B.

Twice the quantity of x produced must not exceed three times the quantity of y produced. Twice x is 2x and three times y is 3y. The phrase "must not exceed" is interpreted as being either equal to it or less than it. Thus 2x can be less than or equal to 3y which is expressed as 2x $ 3y.

2x > 3y means that twice the quantity of x exceeds three times the quantity of y, which is not the given information.

3x ;: 2y means that three times the quantity of x is greater than or equal to twice the quantity of y.

3x < 2y means that three times the quantity of x is less than twice the quantity of y.

Answer to TY 3

The correct option is A.

The feasible area gives you the maximum possible profit given the constraints. All other options are irrelevant.

Answer to TY 4

The correct option is A.

The line passes through co-ordinates x = 6 on the X - axis and y = 4 on the Y - axis. The co-ordinates at the point x = 6 are (6, 0) and at the point y = 4 are (0, 4). We will have to check which of the given equations satisfies these values.

(a) 2x +3y <= 12

Converting inequality into equality
2x + 3y = 12

At (6, 0) - 2(6) + 3(0) = 12
At (0, 4) – 4(0) + 3(4) = 12

As a result, this equation is satisfied by the co-ordinates on the graph. Since all the other options are bound to be wrong we do not need to solve them. There is not even an option which says that two of the options are correct. However, in the examination you might not find the first option to be the correct one. Therefore you will have to do this exercise for all the options in order to find the correct one.

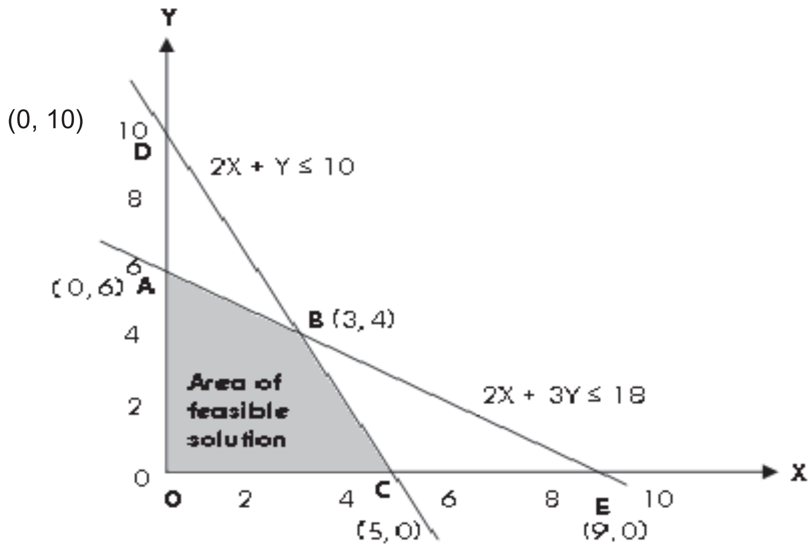
Answer to TY 5

The correct option is A.

This LP problem has two variables, X and Y, and can therefore be solved by using a graph and trial-and-error approach. Two steps are followed. First, plot each constraint so that the area of feasible solutions can be identified. To plot a given constraint, change it from an inequality to an equality, then let Y = 0 and solve for X, and let X = 0 and solve for Y

When Y = 0, 2X + 3x0 = 18
X = 9
When X = 0, 2x0 + 3Y = 18
Y = 6

Applying the same procedures to the Machine 2 constraint, X = 5 when Y = 0, and Y= 10 when X = 0. Plotting these constraints:



The second step is to find the optimal solution, OABC which is the corner points of the area of feasible solutions that maximises total contribution margin when objective function is Z = 4X + 3Y:

Co-ordinates of point B are calculated as follows

The 2X + 3Y :$ 18 line intersected the 2X + Y :$ 10 line. Therefore, by solving the equations, we get:
X = 3 and Y = 4

Point	Corner	X	Y	Total Contribution Margin (Tshs'000)
C	(5,0)	5	0	4(5) + 3(0) = 20
B	(3,4)	3	4	4(3) + 3(4) = 24
A	(0,6)	0	6	4(0) + 3(6) = 18
O	(0,0)	0	0	4(0) + 3(0) = 0

Therefore, the maximum contribution margin is Tshs24,000, which is at the corner point of (3, 4).

Answer to TY 6

Let the number of units of P to be produced be ‘x’ and the units of Q to be produced be ‘y’.

The profit per unit of P is given as Tshs4,000 per unit, and for Q is given as Tshs3,000 per unit. The objective will therefore be:

Maximise 4,000x + 3,000y

The constraints are given as

	Product P (hr / unit)	Product Q (hr / unit)	Available capacity	
Process I	8	8	560	Constraint 1
Process II	2	4	200	Constraint 2
Process III	5	2	320	Constraint 3

The constraints can be converted into inequalities as follows

Constraint 1	8x + 8y $ 560
Constraint 2	2x + 4y $ 200
Constraint 3	5x + 2y $ 320

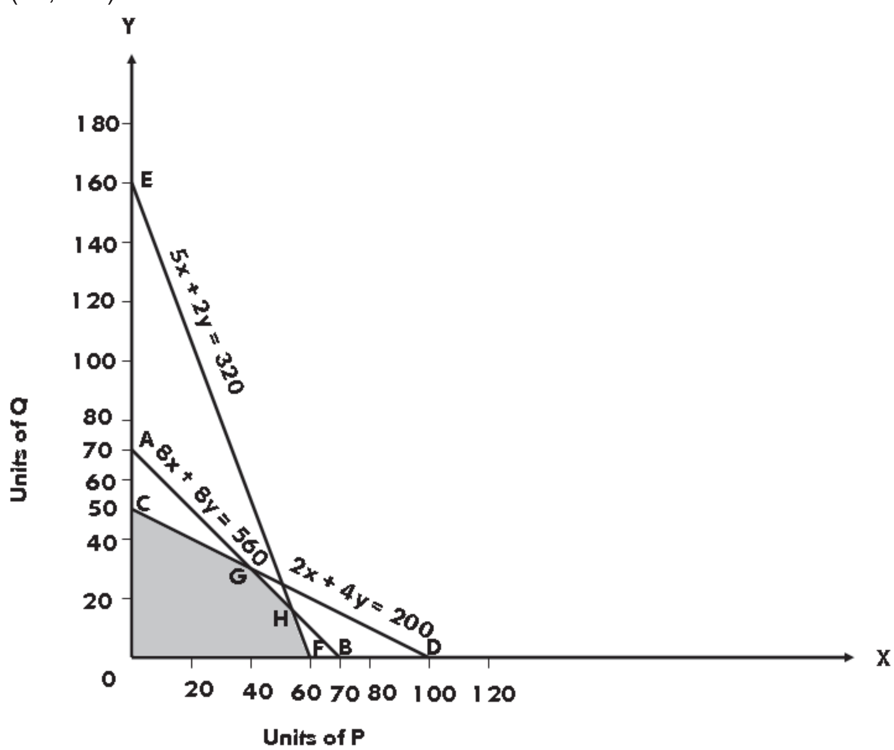
To plot the lines we will first convert these inequalities into equalities.

Constraint 1 – 8x+8y = 560
Putting x = 0 we get 8y = 560 therefore y = 560/8 = 70
Putting y = 0 we get 8x = 560 therefore x = 560/8 = 70

The co-ordinate - (70, 70)

Constraint 2 – 2x+4y = 200
Putting x = 0 we get 4y = 200 therefore y = 200/4 = 50
Putting y = 0 we get 2x = 200 therefore x = 200/2 = 100
The co-ordinate - (100, 50)

Constraint 3 – 5x+2y = 320
Putting x = 0 we get 2y = 320 therefore y = 320/2 = 160
Putting y = 0 we get 5x = 320 therefore x = 320/5 = 64
The co-ordinate - (64, 160)



The feasible area is the pentagon (a shape with five sides) OCGHF. To find the co-ordinates of the points G and H we will have to solve the equations for the lines of which these are the points of intersection.

The three equations we have are:

$8x + 8y = 560$ ------------- (1)
$2x + 4y = 200$ ------------- (2)
$5x + 2y = 320$ ------------- (3)

Solving equations (1) and (2) for coordinate G we have

Multiplying equation (2) by 4 and subtracting equation (1) from it we get

$$\begin{array}{rl} & 8x + 16y = 800 \\ (-) & \underline{8x + 8y = 560} \\ & 8y = 240 \\ & y = 240/8 \\ & y = 30 \end{array}$$

Substituting y = 30 in equation (1) we get

$$\begin{aligned} 8x + 8(30) &= 560 \\ 8x &= 320 \\ x &= 320/8 \\ x &= 40 \end{aligned}$$

The co-ordinates for G are (40, 30)

To find the co-ordinates for point H we need to solve equations (1) and (3)
Multiplying equation (3) by 4 and subtracting equation (1) from it we have

$$\begin{array}{rl} & 20x + 8y = 1{,}280 \\ (-) & \underline{8x + 8y = 560} \\ & 12x = 720 \\ & X = 60 \end{array}$$

Substituting x = 60 in equation (1) we get

$$\begin{aligned} 8(60) + 8y &= 560 \\ 8y &= 80 \\ y &= 80/8 \\ y &= 10 \end{aligned}$$

The co-ordinates of H are (60, 10)

We will now check the profits at each of the co-ordinates at the edge of the feasible area.
(Amounts in Tshs'000)

At C (0, 50): $4x + 3y = 4(0) + 3(50) = 150$
At G (40, 30): $4x + 3y = 4(40) + 3(30) = 250$
At H (60, 10): $4x + 3y = 4(60) + 3(10) = 270$
At F (60, 0): $4x + 3y = 4(60) + 3(0) = 240$

As discussed in the Study Guide the most feasible solution lies only at the edge of the feasible area. We will therefore choose point H (60, 10) where the profit is the highest; this means that we will produce 60 units of product P and 10 units of product Q to earn the maximum profit of Tshs270,000.

Answer to TY 7

Answer

(a) The objective is to maximise the total profit from selling belts A and B which are produced under time, leather and fancy buckle constraints.

(b) Decision variables:
Let x_1 be the number of units of belt A produced
x_2 be the number of units of belt B produced

(c) Objective function:

Total profit Z = $(40{,}000x_1 + 30{,}000x_2)$ must be maximised. Therefore, maximise $Z = 40{,}000x_1 + 30{,}000x_2$

(d) Constraints:

(i) Time constraint: $2x_1 + x_2 \$ 1000$
(ii) Belts' production constraint: $x1 + x2 \$ 800$
(iii) Fancy buckle constraint for belt A: $x1 \$ 400$
(iv) Buckle constraint for belt B: $x2 \$ 700$

(e) Therefore, the LPP format will be –

Maximise	$Z = 40{,}000x_1 + 30{,}000x_2$	(Profit maximisation)
	Subject to	
	$2x_1 + x_2 \$ 1{,}000$	(Time constraint)
	$x_1 + x_2 \$ 800$	(Leather constraint)
	$x_1 \$ 400$	(Fancy buckle constraint)
	$x_2 \$ 700$	(Buckle constraint)
	$x_1 ;: 0, x_2 ;: 0$	(Non-negativity constraint)

(f) The primal equations convert into dual

Therefore, the LPP format is:

Minimise	$C = 1{,}000m_1 + 800m_2 + 400m_3 + 700m_4$	(Cost minimisation)
Subject to		
	$2m_1 + m_2 + m_3 ;: 40{,}000$	
	$m_1 + m_2 + m_4 ;: 30{,}000$	
	$m_1, m_2, m_3, m_4 ;: 0$ (m1, m2, m3 and m4 are not defined)	

Answer to TY 8

Solving the linear programming problem by simultaneous equation method

The constraints are converted into the following equations:

$15x + 24y = 600$………… (1)
$30x + 24y = 720$………… (2)

By subtracting equation (1) from (2), we get:
X = 8

By putting the value of X = 8 in equation (1) we get:
Y = 20

Therefore, the optimum solution for the above constraints is X = 8 and Y = 20.
The shadow price for assembly time is calculated as follows:
The constraints of assembling time become

$30x + 24y \$ 721$ (1)
$15x + 24y \$ 600$ (2)

Shadow price can be calculated by adding one additional unit to the scare resources.

By solving it we get x = 8.067 and y = 19.96

(Amounts in Tshs'000)

Contribution at (8,20) = 1.5(8) + 1.8(20) = Tshs48
Contribution at (8.067, 19.96) = 1.5(8.067) + 1.8(19.96) = Tshs48.03

Therefore, the shadow price per assembling hour is Tshs30.

Answer to TY 9

The optimal solution can be obtained by changing the relevant constraints from inequalities to equalities and solving the resulting simultaneous equations:

2X + 3Y = 18 ………………….equation (i)
2X + Y = 10 ………………….equation (ii)
Subtracting the (ii) equation from the (i),

2Y = 8
Y = 4

By putting Y = 4 into the (i) equation,

2X + 3(4) = 18
2X = 6
X = 3

At this level of production, optimum profit will be:
(Tshs4,000X+3,000Y) = (Tshs4,000 x 3 + 3,000 x 4) = (Tshs12,000+12,000) = Tshs24,000

Answer to TY 10

By substituting the objective variable in the constraints with the optimal values (i.e. 3 units of the first product and 4 units of the second), the following situations arise:

For machine M1:
2X + 3Y $ 18
2 x (3) + 3 x (4) $ 18
6 + 12 = 18

This signifies that the full capacity of machine M1 was utilised in maximising the contribution. For this resource, i.e. machine M1 time, there is no slack.

For machine M 2:
2X + Y $ 10
2 x (3) + 4 $ 10
6 + 4 = 10

This signifies that the full capacity of machine M2 was utilised in maximising contribution. For this resource, i.e. machine M2 time, there is no slack.

Self Examination Questions

Question 1

The most profitable combination of x and y for a linear programming problem is likely to lie:

A On the X-axis
B In the middle of the feasible area
C At the co-ordinates (0, 0)
D At the boundaries of the feasible area

Question 2

The labour hours are limited to 3,000 for the year, with the per unit requirement of hours for the two products as 2 hours per unit of x and 3 hours per unit of y. The machine can manufacture only 5,000 units in a year. Assuming that the units to be manufactured are denoted by 'x' and 'y', the constraints will be given as:

A 3x + 2y $ 3,000 and x + y $ 5,000
B 3x + 2y ;: 3,000 and x – y $ 5,000
C 3x – 2y $ 3,000 and x + y $ 5,000
D 2x + 3y $ 3,000 and x + y $ 5,000

Question 3

A technique of using graphical method of solving linear programming is possible only if:

A The organisation manufactures only two products.
B The organisation manufactures more than two products.
C The organisation manufactures only one product.
D All of the above.

Question 4

Linear programming is a:

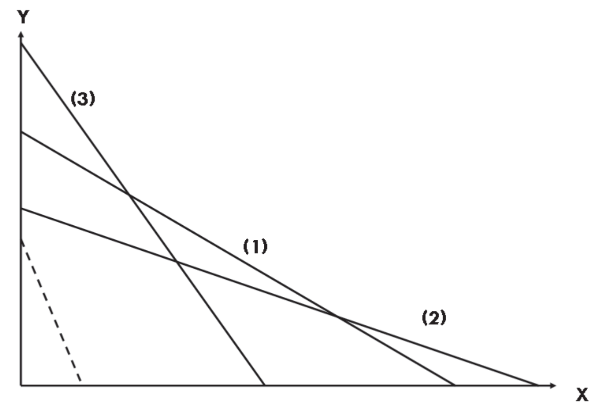
A Mathematical technique
B Statistical technique
C Scientific technique
D Historical technique

Question 5

Grape Wine Inc produces blended whisky and malt whisky every year. Select the inequality which would prove that the production of blended whisky is at least half the production of malt whisky. Assume that the quantity of blended whisky produced throughout the year is X cartons and the quantity of malt whisky is Y cartons.

A ½ X $ Y
B X ;: ½ Y
C X $ ½ Y
D ½ X ;: Y

Question 6



The dotted line depicts the function for maximising contribution. There are three constraints to this function, (1) (2) and (3), which state that the contribution is less than or equal to the required contribution.

State the point of intersection where maximum contribution is achieved.

A (1) and (2)
B (3) and the x- axis
C (3) and (2)
D (3) and (1)

Question 7

A company makes three products P_1, P_2, P_3 which go through three departments – assembly, packaging and storage. The hours of department time required by each of the products, the hours available in each of the departments and the profit contribution of each of the products are given in the following table.

Product	Time required per unit hours)			Contribution Tshs'000 per unit)
	Assembly	Packaging	Storage	
P_1	6	5	9	3
P_2	3	2	7	5
P_3	5	5	10	10
Hours available	210	240	260	

The marketing department of the company indicates that the sales potential for products P_1 and P_2 is unlimited, but for P_3 it is not more than 25 units.

Required:

Formulate the linear programming problem and also find the dual formulation of the LPP.

Question 8

Solve the following problem by the graphical method and by the simultaneous equation method:
Minimise C = 5A + 4B
Subject to the constraints:
4A + B ;: 40
2A + 3B ;: 90
A ;: 0, B ;: 0

Question 9

A firm assembles and sells two different types of computers, A and B, using four resources. The production process can be described as follows:

Resources	Capacity per month
Hard Disk	480 Type A units or 300 Type B units or any linear combination of the two.
Type A Processor	210 Type A units
Type B Motherboard	270 Type B units
Final assembly resource	240 Type A units or 420 Type B units or any linear combination of the two.

Type A units bring in a profit of Tshs18,000 each and type B units, Tshs12,000 each. What should be the optimum product mix?

Answers to Self Examination Questions

Answer to SEQ 1

The correct option is D.

The feasible area is the area which depicts all possible combinations of x and y that yield profits within the given constraints. The boundaries of this feasible area are the maximum possible values of x and y which can yield profit. Beyond this point any combinations of x and y will be impossible due to the constraints.

Answer to SEQ 2

The correct option is D.

The hours required per unit of 'x' are 2, and per unit of 'y' are 3. As such, the constraint will be formulated as:
2x + 3y $ 3,000

The machine cannot produce more than 5,000 units and therefore the production cannot exceed 5,000 units.

This constraint will be:

x + y $ 5,000

All the other options are incorrect as they do not depict each of the constraints properly.

Answer to SEQ 3

The correct option is A.

Linear programming involves two variables which can be plotted on a two axis graph consisting of an X - axis and a Y - axis. Therefore it applies only when an organisation manufactures two products. In any of the other cases given it will not apply.

Answer to SEQ 4

The correct option is A.

Linear programming involves formulating linear equations, which is a mathematical technique. All the other options are automatically wrong as they deal with different subjects.

Answer to SEQ 5

The correct option is B.

This shows that the blended whisky produced is at least half the quantity of the malt whisky produced according to the requirement.

Option A states that the malt whisky produced should be greater than ½ the quantity of blended whisky produced.

Option C states that the blended whisky produced is less than half of the production of the malt whisky.

Option D states that the malt whisky produced should be less than half the quantity of the blended whisky produced.

Answer to SEQ 6

The correct option is B.

This is because the line maximising contribution may move up to this point of the intersection in the feasible area, as this is the farthest point.

Answer to SEQ 7

(a) The objective is to maximise the total profit by selling units of P_1, P_2 and P_3 which are produced under the given production time constraints and sales constraint.

(b) Decision variables:

Let p_1 be the number of units of product P_1 produced
p_2 be the number of units of product P_2 produced
p_3 be the number of units of product P_3 produced

(c) Objective Function: Total profit to be maximised.
Therefore, maximise $Z = 3p_1 + 5p_2 + 10p_3$

(d) Constraints

(i) Assembly department constraint: Total time required for assembly i.e. $6p_1 + 3p_2 + 5p_3$) must be less than or equal to the time available i.e. 210 Hours

Therefore, $6p_1 + 3p_2 + 5p_3$ $ 210

(ii) In a similar manner, packaging department constraint:
$5p_1 + 2p_2 + 5p_3 \$ 240$

(iii) Similarly, store department constraint:
$9p_1 + 7p_2 + 10p_3 \$ 260$

(iv) Sales Constraint for product P_3: the sale is not likely to exceed 25 units. Hence, $p_3 \$ 25$

(e) The LPP formulation will be:

Objective function: Maximise profit, $Z = 3p_1 + 5p_2 + 10p_3$

Subject to

Assembly constraint:	$6p_1 + 3p_2 + 5p_3 \$ 210$
Packaging constraint:	$5p_1 + 2p_2 + 5p_3 \$ 240$
Storage constraint:	$9p_1 + 7p_2 + 10p_3 \$ 260$
Sales constraint:	$p_3 \$ 25$
Non-negativity condition:	p_1, p_2, p_3 ;: 0

(f) The primal problem may be converted into dual as follows:

Objective function: Minimise cost, $C = 210m_1 + 240m_2 + 260m_3 + 25m_4$

Subject to:

$6m_1 + 5m_2 + 9m_3$;: 3
$3m_1 + 2m_2 + 7m_3$;: 5
$5m_1 + 5m_2 + 10m_3+m_4$;: 10

Non-negativity condition: m_1, m_2, m_3, m_4 ;: 0
(m1, m2, m3, m4 is not defined)

Answer to SEQ 8

Solving the linear programming problem by graphical method

Step 1

(a) Find out the points of intersection of the equation 4A + B ;: 40 with the axes.

First this inequality is converted into the equation 4A + B = 40

So as to obtain the points of intersection with the axes

Put A = 0,
4A + B = 40
4 x (0) + B = 40

$\boxed{B = 40}$

Again,
Put B = 0,
4A + B = 40
4A + 0 = 40

$\boxed{A = 10}$

Alternatively, by expressing the equation 4A + B = 40 into intercept form, we get, $\frac{A}{10} + \frac{B}{40} = 1$

Accordingly, we can directly infer that the straight line intersects the horizontal axis at (10,0) and the vertical axis at (0,40) points.

(b) Find out the points of intersection of the equation 2A + 3B ;: 90 with the axes

First this inequality is converted into the equation 2A + 3B = 90
So as to obtain the points of intersection with the axes,

Put A = 0,
2A + 3B = 90
2 x (0) + 3B = 90

$\boxed{B = 30}$

Again,
Put B = 0,
2A + 3 x (0) = 90
2A + 0 = 90

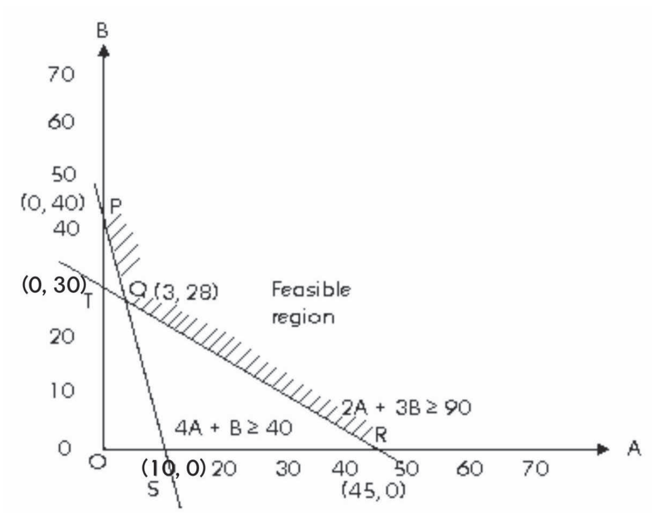
$\boxed{A = 45}$

Alternatively, by expressing the equation 2A + 3B = 90 into intercept form, we get, $\frac{A}{45} + \frac{B}{30} = 1$

Accordingly, we can directly know that the straight line intersects the horizontal axis at (45, 0) and the vertical axis at (0,30) points.

Step 2

Graphical presentation of Linear Programming problem



Step 3

Find out co-ordinate corner points of the feasible region

(a) Co-ordinate of point Q

At the point Q, the 4A + B ;: 40 equation line intersected the 2A + 3B ;: 90 equation line. Therefore, calculating the co-ordinates of point Q by solving these two equations under the simultaneous equation method, we get

4A + B = 40 ……………......equation (i)
2A + 3B = 90 ……………….. equation (ii) multiplied by 2

$\boxed{\text{4A + 6B = 180 …………….. equation (iii)}}$

4A + 6B = 180 ………….........equation (iii)
-) 4A + B = 40 ……………....equation (i)

$\boxed{\text{5B = 140 ……………… Equation (iii) – (i)}}$

$\boxed{B = 28}$

By putting the value of B in equation (i), calculate the value of A

4A + B = 40
4A = 12

$\boxed{A = 3}$

Co-ordinates of corner points of the PQR feasible region

Point	(X, Y)
P	(0, 40)
Q	(3, 28)
R	(45, 0)

Step 4

Find out the optimum minimum) cost by putting the value of the co-ordinates of the corner points of the feasible region into the objective function C = 5A + 4B

Corner point	Co-ordinates of corner points	Objective function C = 5A + 4B (Tshs'000)	Value (Tshs'000)
P	(0, 40)	C (P) = 5 x (0) + 4 x (40)	160
Q	(3, 28)	C (Q) = 5 x (3) + 4 x (28)	127
R	(45, 0)	C (R) = 5 x (45) + 4 x(0)	225

Out of the three alternatives, the second one i.e. for point (Q), gives the minimum value of Tshs127,000 indicating that the optimum value of the objective function is Tshs127,000.

Solving the linear programming problem by simultaneous equation method

For solving the problem under the simultaneous equation method, the constraints are converted into the following equations:

4A + B = 40 ………..... (i)
2A + 3B = 90 ………….. (ii)

By multiplying both the sides of the equation (ii) by 2,

We get,
4A + 6B = 180 …………. (iii)

Deducting equation (i) from this, we will get,

5B = 180 – 40
Or, 5B = 140
Or, B = 28

By substituting the value of B in equation (i), we get

4A + 28 = 40
Or, 4A = 12
Or, A = 3

Therefore, the optimum value of the objective function would be 5 x (3) + 4 +(x) (28) = 127

Answer to SEQ 9

Step 1 (define x and y)

Let X = type A processor and Y = type B motherboard

Appropriate mathematical formulation of the above problem

Maximise Z = 18000X + 12000Y

Subject to constraint: 1/480) X + 1/300) Y $ 1 or 5X + 8Y $ 2400
1/240) X + 1/420) Y $ 1 or 7X + 4Y $ 1680
X $ 210 and Y $ 270
X ;: 0, Y ;: 0

Step 2

(a) Find out the points of intersection of the equation 5X + 8Y $ 2,400 with the axes.

First this inequality is converted into the equation 5X + 8Y = 2,400
So as to obtain the points of intersection with the axes,

Put X = 0,
5X + 8Y = 2,400
5 x (0) + 8 x (Y) = 2,400

$\boxed{Y = 300}$

Again,

Put Y = 0,
5X + 8Y = 2,400
5X + 8 x (0) = 2,400

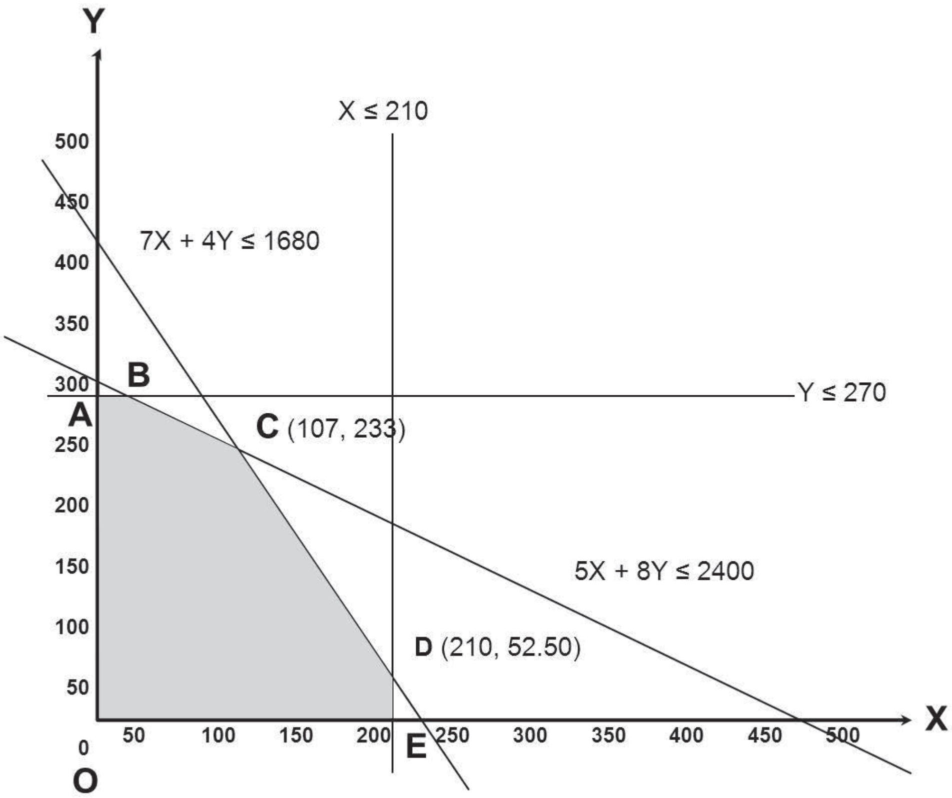
$\boxed{X = 480}$

(b) Find out the points of intersection of the equation 7X + 4Y $1,680 with the axes.
First this inequality is converted in the equation 7X + 4Y = 1,680.
So as to obtain the points of intercepts with the axes,
Put X = 0 in
7X + 4Y = 1,680
7 x (0) + 4Y = 1,680
$\boxed{Y = 420}$

Again,
Put Y = 0 in
7X + 4Y = 1,680
7X + 4 x (0) = 1,680
$\boxed{X = 240}$

Step 3

Graphical presentation of the Linear Programming problem



(a) Co-ordinates of point B,

Y $ 270 line intersected to 5X + 8Y $ 2,400 line. Therefore, by putting the value of Y i.e. 270) in the equation 5X + 8Y = 2,400, we get X = 48. Therefore, the co-ordinates of point B are 48, 270)

(b) Co-ordinates of point C,

Point C, at which 7X + 4Y $ 1,680 equation line intersected the 5X + 8Y $ 2,400 the equation line. Therefore, calculate the co-ordinates of point C by solving these two equations under the simultaneous equation method

7X + 4Y = 1,680 …………………………..equation (i) multiplied by 2
14X + 8Y = 3,360 …………………………..equation (iii)

5X + 8Y $ 2,400 ……………………………Equation (ii)
14X + 8Y = 3,360 ………………………. Equation (iii)
5X + 8Y = 2,400…………………………..Equation (ii)
9X = 960………………………………..Equation (iii) – (ii)

X = 320/3 or 107 approx.

By putting the value of X = 320/3) in equation II), calculate the value of Y.

5X + 8Y = 2,400…………………………..Equation iv)
5 x (320/3) + 8Y = 2400
(1600/3) + 8Y = 2,400

Y = 700/3 or 233 approx.

Therefore, the co-ordinates of point C are (320/3, 700/3) or (107, 233) approx

(c) Co-ordinates of point D,

X $ 210 line intersected the 7X + 4Y $ 1,680 line. Therefore, by putting the value of X in the equation 7X + 4Y = 1,680, we get Y = 52.5. Therefore, the co-ordinates of point Dare (210, 52.5)

Point	(X, Y)
O	(0, 0)
A	(0, 270)
B	(48, 270)
C	(107, 233)
D	(210, 52.5)
E	(210, 0)

Step 4

Find out the optimum profit by putting the value of the co-ordinates of the corner points of the feasible region in the objective function Z = 18,000X + 12,000Y

Corner point	Co-ordinate of corner points	Objective function Z = 18X + 12Y (Tshs'000)	Value (Tshs'000)
O	(0, 0)	Z (O) = 18 x(0) + 12 x (0)	0
A	(0, 270)	Z (A) = 18 x (0) + 12 x (270)	3,240
B	(48, 270)	Z (B) = 18 x (48) + 12 x (270)	4,104
C	(107, 233)	Z (C) = 18 x (107) + 12 x (233)	4,722
D	(210, 52.5)	Z (D) = 18 x (210) + 12 x (52.5)	4,410
E	(210, 0)	Z (E) = 18 x (210) + 12 x (0)	3,780

Out of the six options, the fourth one, i.e. for point C), gives the maximum profit of Tshs4,722,000 indicating the optimum profit for the business.

9

TRANSPORTATION AND ASSIGNMENT MODELS

Get Through Intro

Corporations today have to be cost effective to be able to compete and sustain in the industry. Given the dynamic nature of economic and other variables affecting a business, companies have to take effective steps not only to control their overall costs, but also to reduce the per unit cost. This is all the more challenging if the demand and supply is widely dispersed and the entity faces the difficult task of keeping the transportation costs at a minimum to maximise profits.

Furthermore, the production activity needs to be broken down and assigned to the best possible resources, while at the same time managing the constraints of time and costs to maximise profit.

This Study Guide will enable you to understand the techniques to solve transportation problems and assignment problems to achieve the objective of maximising profits. Application of the techniques discussed in this Study Guide will help you to manage your business more efficiently.

Learning Outcomes

a) Understand the meaning of transportation problems and transportation models and their application in business activities.
b) Use of linear programming models to formulate a transportation problem.
c) Understand balanced transportation problems and unbalanced transportation problems.
d) Solve a transportation problem using North West Corner method, Minimum Cost Method and Vogel Approximation Method.
e) Use stepping stone approach for testing.
f) Apply the concept of transportation problem in accounting and business situations.
g) Solve an assignment problem by allocation that will produce optimal solutions.
h) Apply the concept of assignment problem to accounting and business situations.

1. Understand the meaning of transportation problems and transportation models and their application in business activities.
[Learning Outcome a]

1.1 Meaning of transportation problem

Definition

Transportation problem is a problem faced in transporting various amounts of homogenous commodities, which are initially manufactured, produced or stored at various locations, to different destinations in such a manner that the total transportation cost is minimised.

In other words, a transportation problem is a particular class of linear programming problems, which is associated with day-to-day activities in our real life and mainly deals with problems associated with logistics planning. It helps in solving problems on distribution and transportation of resources from one place to another. The objective of a transportation problem like a linear programming problem is to satisfy the demand at various destinations from the supply available at various sources. However, this is subject to constraints, in the form of maximum quantity available at any source (e.g. units available at the factory) and the maximum quantity that can be received at any destination (e.g. warehouse), at the minimum transportation cost possible.

Important

Linear programming is a technique used to find a solution for optimising a given objective under certain constraints.

Example

A motor spare parts seller has warehouses located at three different regions: Dar-es-Salam, Arusha and Dodoma in Tanzania. This company produces motor parts at these locations with capacities 6000, 5000 and 4000 units per week respectively. The products are sold through four distributors in the country, located at Kariakoo (DSM), USA river (Arusha) and Msalato (Dodoma). The weekly demand for the spare parts is 5000, 4000, 2000 and 4000 units at USA River, Kisasa, Kariakoo, and Ubungo respectively. The cost of transportation per unit from each of the plants to the distributors varies. The company would like to determine the number of units to be shipped from each plant to each particular distributor to satisfy the demand.

This is a transportation problem.

1.2 Transportation models

The transportation model is concerned with selecting the routes between supply and demand points in order to minimise costs of transportation subject to constraints of supply at any supply point and constraints of demand at any demand point. These are a type of linear programming models useful to solve transportation problems by satisfying the demand at various destinations from the supply constraints at the minimum transportation cost possible.

The transportation model can also be used in making location decisions. The model helps in locating a new facility, a manufacturing plant or an office, when two or more number of locations are under consideration. The total transportation cost, distribution cost or shipping cost and production costs can be minimised by applying the model.

Example

Assume a company has 4 manufacturing plants with different capacity levels, and 5 regional distribution centres. Therefore, 4 x 5 = 20 routes are possible. Given the transportation costs per load of each of 20 routes between the manufacturing (supply) plants and the regional distribution (demand) centres, and supply and demand constraints, how many loads can be transported through different routes so as to minimise transportation costs? The answer to this question is obtained easily through the transportation model.

1.3 Application of transportation model in business activities

Listed below are certain applications of the transportation models in business activities

1. To decide the transportation of new materials from various centres to different manufacturing plants. In the case of a multi-plant company, this will help minimise (transport) costs and hence maximise profits.

Example

A manufacturing company has five workshops situated all over Tanzania. The company manufactures furniture in these workshops. It is required to procure three different varieties of wood for its productions from three different locations and the demand for the wood is different for all the five workshops. The company wants to determine the most cost effective way of transporting wood from the three locations to its workshops.

The transportation model can provide a solution to the company's problems.

---

2. To decide the transportation of finished goods from different manufacturing plants to the different distribution centres. For a multi-plant-multi-market company, this will help minimise (transport) costs and hence maximise profits.

Example

A soft drink manufacturing firm has 3 plants located in 3 different cities. The total production is absorbed by 120 retail shops in 50 different cities. We want to determine the transportation schedule that minimises the total cost of transporting soft drinks from various plants to various retail shops.

---

3. To make location decisions for plants as well as warehouses.

Example

XYZ Ltd is a trading company which deals in a variety of products. The company buys products from five different manufacturing companies whose plants are located at Tanga, Kasulu, Moshi, Songea and Iringa. It has its three warehouses located at Babati, Handeni and Mafinga. The company wants to know the most cost effective way in terms of transportation of goods from various plants to the various warehouses.

---

4. To help in locating a new facility, a manufacturing plant or an office when two or more locations are under consideration.

Test Yourself 1

Which of the following is the main purpose of a Transportation model?

A Transport goods to the supply points
B Allocate goods to customers
C Distribute a single good from various sources to different destinations at minimum total cost
D Sell goods to consumers

---

2. Use of linear programming models to formulate a transportation problem. [Learning Outcome b]

2.1 Programming of transportation model

A transportation problem applies to situations where a single commodity is to be transported from various sources of supply (origins) to various demand points (destinations) at minimum cost.

Often, businesses have a number of plants where goods are manufactured and a number of warehouses where goods are to be transported. There is a different per unit transportation cost from each plant to each warehouse. The concern of these businesses would be to transport goods from the plants to the warehouses at the lowest possible cost.

This can be done by formulating a linear programming model in the following way:

(a) For a given product P,
assume
There are 'm' number of supply sources (e.g. manufacturing plants) termed $S_1, S_2, S_3, \ldots, S_m$
Each supply source, Si, produces ai units of the product

There are n number of destinations (e.g. warehouses) termed $D_1, D_2, D_3, \ldots, D_n$
Each destination Dj, requires bj units of the product

(b) Let c_{ij} be the cost of transporting one unit from S_i to D_j
Let x_{ij} represents the number of units transported from S_i to D_j

(c) Now the problem is to determine the transportation schedule which minimises the total transportation cost termed z, for satisfying the above supply and demand requirements.

The above transportation problem can be stated mathematically as a linear programming model as below:

$$z = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} c_{ij}$$

where i = 1, 2, 3, …, m & j = 1, 2, 3, …, n

Constraints in a transportation problem

In any business unit, the production capacity, i.e. units available to supply (ai) of a given plant, is limited and at the same time the units that can be stored, i.e. demanded (bj) at any warehouse, is also limited. Here, we need to decide how much quantity is to be transported from each source to each destination so that the demands are also satisfied and the units transported also do not exceed the supply and demand capacities of any source or destination respectively.

Therefore, the above formula for total cost is also subject to supply and demand constraints.

The formulae for supply and demand constraints are given below:

Supply constraints

$$\sum_{j=1}^{n} x_{ij} = a_i$$

Demand constraints

$$\sum_{j=1}^{n} x_{ij} = bj$$

and x_{ij} (number of units to be transported) ≥ 0 , for all sources and destinations.

To summarise,

m	Represents number of sources of supply i.e. $S_1, S_2, S_3 \ldots\ldots\ldots S_m$
n	Represents number of destinations (Demand) centres i.e. $D_1, D_2, D_3, \ldots\ldots.D_n$
i	Represents any given source
j	Represents any given destination
a_i	Represents units of the commodity available at source i to be transported.
b_j	Represents units of the commodity needed at destination j
c_{ij}	Represents cost of transporting one unit of the commodity from ith source to jth destination
x_{ij}	Represents the units to be transported from ith source to jth destination
Z	Represents total cost of transporting all the units from all the sources to the destinations.

2.2 Tabular Representation of transportation model

Based on the above information, the transportation problem can also be represented in a tabular form as shown below:

Let:
C_{ij} be the cost of transporting a unit of the product from i^{th} origin to j^{th} destination, wherein i = 1,2,3 …..n and j = 1,2,3, …..m
ai be the quantity of the commodity available at source i,
bj be the quantity of the commodity needed at destination j, and
xij be the quantity transported from i^{th} source to j^{th} destination

	Destination				Supply available at each source
Source	To / From	D_1	D_2	D_n	A_i
	S_1	C_{11} X_{11}	C_{12} X_{12}	D_{1n}	a_1
	S_2	C_{21} $C2_1$	C_{22} X_{22}	$C2_n$	a_2
	Up to S_m	C_{m1} X_{m1}	C_{m2} X_{m2}	C_{mn}	A_m
Demand at each destination	B_j	b_1	b_2	B_n	Here the total demand is equal to total supply

In the table given above,

- the cost of transporting a unit from source S_1 to destination D_1 is C_{11}
- the number of units which can be transported from S_1 to D_1 is X_{11}

Here we have assumed that the quantity available at source S_1 (i.e. ai) is a_1 and the quantity demanded at destination D_1 (i.e. b_j) is b_1. Therefore, C_{ij} is C_{11} and X_{ij} is X_{11}

Similarly, the cost of transportation and the number of units to be transported is shown for other routes also.

Now let us understand the formulation of the transportation model with help of an example:

Example

Sodexo Inc is a company, manufacturing footwear, and has production facilities at Tanga, Dodoma and Arusha. The production capacities at Tanga, Dodoma and Arusha are 6,000, 5,000 and 4,000 units per week respectively.

The textile unit distributes its shoes through four of its wholesale distributors situated at four locations: Tabora, Kigoma, Singida and Kahama. The weekly demand of the distributors is 3,000, 2,000, 2,000 and 4,000 units for Tabora, Kigoma, Singida and Kahama respectively. The cost of transportation per unit varies between different supply points and destination points.

The supply, demand and transportation costs are summarised as follows:

Manufacturing Capacities

Supply	Footwear Manufacturing (units)	Monthly Productions (units)
1	Tanga	6,000
2	Dodoma	5,000
3	Arusha	4,000

Demand Requirements

Demand	Showroom	Monthly Demand (units)
1	Tabora	3,000
2	Kigoma	2,000
3	Singida	2,000
4	Kahama	4,000

Transportation Cost per unit (Tshs'000)

Supply	Destination			
	Tabora	Kigoma	Singida	Kahama
	Tshs'000	Tshs'000	Tshs'000	Tshs'000
Tanga	5	6	9	7
Dodoma	7	8	2	4
Arusha	5	3	5	3

The management of Sodexo Inc wants to know the number of units to be transported from each location to each distributor so that the total transportation cost is minimised.

Required:

Work out the linear programming model for the above transportation problem.

Answer

A linear programming model can be used to solve the transportation problem.

As in the tabular representation shown above, let

X_{11} be number of units shipped from source 1 (Tanga) to destination 1 (Tabora).
X_{12} be number of units shipped from source 1 (Tanga) to destination 2 (Kigoma).
X_{13} be number of units shipped from source 1 (Tanga) to destination 3 (Singida).
X_{14} be number of units shipped from source 1 (Tanga) to destination 4 (Kahama)
and so on.

X_{ij} = number of units shipped from source i to destination j,

Continued on the next page

where i = 1, 2, …, m and j = 1, 2, …, n

Using the cost data table given above, the following equations can be arrived at:

Transportation cost for units shipped from Tanga = (5 x X_{11}) + (6 x X_{12}) + (9 x X_{13}) + (7 x X_{14})
Transportation cost for units shipped from Dodoma = (7 x X_{21}) + (8 x X_{22}) + (2 x X_{23}) + (4 x X_{24})
Transportation cost for units shipped from Arusha = (5 x X_{31}) + (3 x X_{32}) + (5 x X_{33}) + (3 x X_{34})

Constraints

In transportation problems, there are supply constraints for each source, and demand constraints for each destination.

Supply constraints

For Tanga, X_{11} + X_{12} + X_{13} + X_{14} =$ 6,000
For Dodoma, X_{21} + X_{22} + X_{23} + X_{24} =$ 5,000
For Arusha, X_{31} + X_{32} + X_{33}+ X_{34} =$ 4,000

Demand constraints

For Tabora, X_{11} + X_{21} + X_{31} = 3,000
For Kigoma, X_{12} + X_{22} + X_{32} = 2,000
For Singida, X13 + X_{23} + X_{33} = 2,000
For Kahama, X_{14} + X_{24} + X_{34} = 4,000

The linear programming model for Sodexo is as follows:

Minimise $Z = 5X_{11} + 6X_{12} + 9X_{13} + 7X_{14} + 7X_{21} + 8X_{22} + 2X_{23} + 4X_{24} + 6X_{31} + 3X_{32} + 5X_{33} + 3X_{34}$

Subject to constraints

$X_{11} + X_{12} + X_{13} + X_{14}$ \$ 6,000 (i)
$X_{21} + X_{22} + X_{23} + X_{24}$ \$ 5,000 (ii)
$X_{31} + X_{32} + X_{33} + X_{34}$ \$ 4,000 (iii)
$X_{11} + X_{21} + X_{31}$ = 3,000 (iv)
$X_{12} + X_{22} + X_{32}$ = 2,000 (v)
$X_{13} + X_{23} + X_{33}$ = 2,000 (vi)
$X_{14} + X_{24} + X_{34}$ = 4,000 (vii)

Where, x_{ij} ;: 0 for i = 1, 2, 3 and j = 1, 2, 3, 4.

Important

The Linear Programming Model shows how to represent a transportation problem in a tabular form and in the form of a mathematical equation. But the actual method of ascertaining the number of units to be transported from each source to each destination so that the total cost is minimised will be discussed in the chapter ahead.

Test Yourself 2

Which of the following is to be minimised in a transportation problem?

A Demand
B Supply
C Total Cost
D Number of units transported

3. Understanding balanced transportation problems and unbalanced transportation problems.

[Learning Outcome c]

3.1 Balanced transportation problem

The supply and demand sides will rarely be equal in any given practical situation. This is because of:

(a) the variations in the production: due to shortage of raw materials, labour problems, improper planning and scheduling from the supplier and

(b) the variations in demand forecast: due to change in customer preference, change in prices and introduction of new products by competitors from the customer end, resulting in demand fluctuations.

When the total supplies of all the sources together are equal to the total demand of all destinations together, the problem is a balanced transportation problem.

Total supply Total demand

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$$

3.2 Unbalanced transportation problem

When the total supply from all the sources is not equal to the total demand at all the destinations, the problem is an unbalanced transportation problem.

Where,

Total supply Total demand

$$\sum_{i=1}^{m} a_i \neq \sum_{j=1}^{n} b_j$$

These unbalanced problems can be easily solved by introducing dummy sources and dummy destinations in the following manner:

Step 1

If the total supply is greater than the total demand, a dummy destination (dummy column) with demand equal to the supply surplus is added.

Step 2

If the total demand is greater than the total supply, a dummy source (dummy row) with supply equal to the demand surplus is added.

Step 3

The unit transportation cost for the dummy column and dummy row are assigned zero values, because no transportation is actually made in case of a dummy source and dummy destination.

3.3 When demand is less than supply

Now, let us learn how to convert an unbalanced problem to a balanced transportation problem when the demand is less than supply, with the help of an example:

Example

The following data regarding demand and supply relates to ABC Ltd

Transportation Cost per unit (Tshs'000)

Source	Destination			Supply (units)
	1	2	3	
	Tshs'000	Tshs'000	Tshs'000	
1	25	45	10	200
2	30	65	15	100
3	15	40	55	400
Demand (units)	200	100	300	

Required:

Convert the above unbalanced transportation problem into a balanced transportation problem.

Answer

For the given problem, the total supply is not equal to the total demand.

Total supply		Total demand
$\sum_{i=1}^{m} a_i = 700$ units	and	$\sum_{j=1}^{n} b_j = 600$ units

The difference between the supply and demand is 100 units.

Total supply		Total demand	
$\sum_{i=1}^{m} a_i = 700$ units	-	$\sum_{j=1}^{n} b_j = 600$ units	100 units

Now let us see how to arrive at the solution using the following steps:

Step 1

We need to add a dummy destination column with a demand of 100 units. The modified table is given below.

Transportation Cost per unit (Tshs'000)

Source	Destination				Supply (units)
	1	2	3	4	
	Tshs'000	Tshs'000	Tshs'000	Tshs'000	
1	25	45	10		200
2	30	65	15		100
3	15	40	55		400
Demand (units)	200	100	300	100	

Step 2

This is not required here, as in this case, the demand is less than the supply.

Continued on the next page

Step 3

We need to assign zero values to the unit transportation cost for the dummy column, because no shipment is actually made to a dummy destination.

The modified table with balanced demand and supply is shown below.

Transportation Cost per unit (Tshs'000)

Source	Destination				Supply (units)
	1	2	3	4	
	Tshs'000	Tshs'000	Tshs'000	Tshs'000	
1	25	45	10	0	200
2	30	65	15	0	100
3	15	40	55	0	400
Demand (units)	200	100	300	100	700/700

3.4 When supply is less than demand

Now, let us learn how to convert an unbalanced problem to a balanced transportation problem when the demand is less than supply, with the help of an example:

Example

The following data regarding demand and supply relates to ABC Ltd

Transportation Cost per unit (Tshs'000)

Source	Destination			Supply (units)
	1	2	3	
	Tshs'000	Tshs'000	Tshs'000	
1	25	45	10	200
2	30	65	15	100
3	15	40	55	200
Demand (units)	200	100	300	

Required:

Convert the above unbalanced transportation problem into a balanced transportation problem.

Answer

For the given problem, the total supply is not equal to the total demand.

Total demand $\sum_{j=1}^{n} b_j = 600$ units and Total supply $\sum_{i=1}^{m} a_i = 500$ units

The difference between the supply and demand is 100 units.

Total demand $\sum_{j=1}^{n} b_i = 700$ units - Total supply $\sum_{i=1}^{m} a_i = 600$ units 100 units

Continued on the next page

Now,

Step 1

We need to add a dummy source row with a supply of 100 units. The modified table is given below.

Transportation Cost per unit (Tshs'000)

Source	Destination			Supply (units)
	1	2	3	
	Tshs'000	Tshs'000	Tshs'000	
1	25	45	10	200
2	30	65	15	100
3	15	40	55	200
4				100
Demand (units)	200	100	300	

Step 2

This is not required here in this case, as the supply is less than the demand.

Step 3

We need to assign zero values to the unit transportation cost for the dummy row, because no shipment is actually made from a dummy source.

The modified table with balanced demand and supply is shown below.

Transportation Cost per unit (Tshs'000)

Source	Destination			Supply (units)
	1	2	3	
	Tshs'000	Tshs'000	Tshs'000	
1	25	45	10	200
2	30	65	15	100
3	15	40	55	200
4	0	0	0	100
Demand (units)	200	100	300	600/600

Test Yourself 3

The supply capacities of plant X, Y, W and Z are as follows:

X = 50,
Y = 70,
W = 60,
Z = 80,

Demand capacities of warehouses are as follows:

A= 10,
B = 70,
C = 90,
D = 110;

Determine the amount of dummy capacity to add to the source or destination.

A 20 (dummy) for supply (source)
B 20 (dummy) for demand (destination)
C 50 (dummy) for supply (source)
D 50 (dummy) for demand (destination)

4. Solve a transportation problem using North West Corner method, Minimum Cost Method and Vogel Approximation Method.
   Use stepping stone approach for testing.
   Apply the concept of transportation problem in accounting and business situations.
   [Learning Outcomes d, e and f]

4.1 Procedure to solve a transportation problem

Step 1: Formulate the problem

Formulate the given problem and set up in a matrix form. Check whether the problem is a balanced or unbalanced transportation problem. If unbalanced, add dummy source (row) or dummy destination (column) as required.

Step 2: Obtain the initial feasible solution

The initial feasible solution can be obtained by the following methods:

1. Northwest Corner Method
2. Least Cost Method
3. Vogel's Approximation Method

Step 3: Check for degeneracy

The number of positive allocations, N must be equal to m + n -1, where m and n are numbers of rows and columns respectively.

Any solution that satisfies the condition N = m + n - 1 is termed "Non-degenerate Basic Feasible solution"; otherwise, it is called "Degenerate solution".

If number of allocations, N = m + n – 1, then degeneracy does not exist. Go to Step 5.

If number of allocations, N f. m + n – 1, then degeneracy does exist. Go to Step 4.

Step 4: Resolving degeneracy

To resolve degeneracy at the initial solution, allocate a small positive quantity 'e' to one or more unoccupied cells that have the lowest transportation costs, so as to make m + n – 1 allocations (i.e., to satisfy the condition N = m + n – 1). The cell chosen for allocating 'e' must be of an independent position. In other words, the allocation of 'e' should avoid a closed loop and should not have a path.

Step 5: Test for optimality

Once an initial solution is obtained, the next step is to test its optimality. An optimal solution is one in which there are no other transportation routes that would reduce the total transportation cost.

Step 6: Calculate the total transportation cost

When all the Cij values are zero or greater than zero, optimality is reached and hence the present allocations are the optimum allocations. Calculate the total transportation cost by assuming the product of allocated units and unit costs.

4.2 Let us understand the following three methods

(i) Northwest Corner Method
(ii) Least Cost Method
(iii) Vogel's Approximation Method

4.3 North West Corner Rule (NWCR) of solving transportation problem

This method is the simplest, but it is the most inefficient method as it has the highest total transportation cost in comparison to all other methods. The main reason that can be attributed to this is that the method does not take into account the cost of transportation for all the possible alternative routes.

The steps to obtain the initial basic feasible solution are as follows:

Step 1 Select the North-west (i.e., upper left) corner cell of the table and allocate the maximum possible units between the supply and demand requirements. During allocation, the transportation cost is completely discarded (not taken into consideration i.e. whether the North-west corner cell has the lowest cost or not does not matter).

Step 2 Delete the row or column which has no values remaining (fully exhausted) for supply or demand.

Step 3 Now, with the new reduced table, again select the North-west corner cell and allocate the available values.

Step 4 Repeat steps (ii) and (iii) until all the supply and demand values are zero.

Step 5 Obtain the initial basic feasible solution.

Now let us understand the way of obtaining the initial basic feasible solution for a transportation problem by North West Corner Rule as discussed above with the help of an example:

Example

The cost of transportation per unit from three sources and four destinations are given in the table below.

Transportation Cost per unit

Source	Destination				Supply (units)
	1	2	3	4	
	Tshs'000	Tshs'000	Tshs'000	Tshs'000	
1	4	2	7	3	250
2	3	7	5	8	450
3	9	4	3	1	500
Demand (units)	200	400	300	300	1,200

Required:

Obtain the initial basic feasible solutions using the North West Corner Method.

Answer

Step 1: Formulate the problem

We need to formulate the problem first and set it up in a matrix form. We then need to check whether the problem is a balanced or unbalanced transportation problem.

Total demand = Total supply

$$\sum_{j=1}^{N} bi = 1,200 \quad = \quad \sum_{i=1}^{m} ai = 1,200$$

The given problem here is a balanced one as the demand and supply are equal, i.e. 1,200 units.

Step 2: Obtain the initial feasible solution

The initial feasible solution here needs to be obtained by the North West Corner Method

Step 1: Select the North-west (i.e., upper left) corner cell of the table and allocate the maximum possible units between the supply and demand requirements. During allocation, the transportation cost is completely discarded (not taken into consideration i.e. whether the North-west corner cell has the lowest cost or not does not matter).

Continued on the next page

In the given matrix, the North-West corner cell is (1,1) and the supply and demand values corresponding to cell (1,1) are 250 and 200 respectively. We need to allocate the maximum possible value to satisfy the demand from the supply. Hence allocate 200 to the cell (1,1) as shown below:

Transportation cost per unit in Tshs ('000)

Source	Destination				Supply (units)
	1	2	3	4	
	Tshs'000	Tshs'000	Tshs'000	Tshs'000	
1	200 4	2	7	3	~~250~~ 50
2	3	7	5	8	450
3	9	4	3	1	500
Demand (units)	~~200~~	400	300	300	

Step 2: Delete the row or column which has no values (fully exhausted) for supply or demand.

In this case, we need to delete the exhausted column 1 which gives a new reduced table as shown below:

Table after deleting Column 1

Source	Destination			Supply (units)
	2	3	4	
	Tshs'000	Tshs'000	Tshs'000	
1	2	7	3	50
2	7	5	8	450
3	4	3	1	500
Demand (units)	400	300	300	

Step 3: Now, with the new reduced table, again select the North-west corner cell and allocate the available values.

In the table given above the North West Corner Cell is (1,2) and the demand and supply values corresponding to it are 400 and 50 respectively. We need to allocate maximum supply values available to the demand. Hence, the supply is fully allocated to demand as shown below.

Transportation cost per unit in Tshs ('000)

Source	Destination			Supply (units)
	2	3	4	
	Tshs'000	Tshs'000	Tshs'000	
1	50 2	7	3	~~50~~
2	7	5	8	450
3	4	3	1	500
Demand (units)	~~400~~ 350	300	300	

Continued on the next page

Step 4: Repeat steps 2 and 3 till all the demand and supply values are zero.

Now, as shown in the table above, the supply of source 1 is exhausted; therefore, we need to delete row 1.

The reduced matrix is shown in the table below.

Table after deleting Row 1

Source	Destination			Supply (units)
	2	3	4	
	Tshs'000	Tshs'000	Tshs'000	
2	7	5	8	450
3	4	3	1	500
Demand (units)	350	300	300	

Furthermore, the North West cell is (2,2) and the corresponding demand and supply values are 350 and 450. Hence, the demand at destination 2 is allocated as shown in the table below.

Source	Destination			Supply (units)
	2	3	4	
	Tshs'000	Tshs'000	Tshs'000	
2	350 7	5	8	~~450~~ 100
3	4	3	1	500
Demand (units)	~~350~~	300	300	

Now, since the demand at destination 2 is fully satisfied, we need to delete column 2 as shown in the table below.

Source	Destination		Supply (units)
	3	4	
	Tshs'000	Tshs'000	
2	5	8	100
3	3	1	500
Demand (units)	300	300	

Again, the North West corner cell is (2,3) and the corresponding demand and supply values are 300 and 100 respectively. Therefore, we allocate the supply of source 2 to it and delete row 2 as shown in the tables below.

Source	Destination		Supply (units)
	3	4	
	Tshs'000	Tshs'000	
2	100 5	8	~~100~~
3	3	1	500
Demand (units)	~~300~~ 200	300	

Continued on the next page

Table after deleting Row 2

Source	Destination		Supply (units)
	3	4	
	Tshs'000	Tshs'000	
3	3	1	500
Demand (units)	200	300	

Now, we have only once source of 500 remaining and two destinations with demand of 200 and 300 remaining; therefore, the available supply is allocated to both demands and we have zero demand and supply values.

Source	Destination		Supply
	3	4	
	Tshs'000	Tshs'000	
3	200 3	300 1	~~500~~
Demand	~~200~~	~~300~~	

Now all the demands and supplies have been allocated.

Step 5: Obtain the initial basic feasible solution.

The initial basic feasible solution using the North-west corner method is shown below:

Transportation cost

= (Tshs4,000 × 200 units) + (Tshs2,000 × 50 units) + (Tshs7,000 × 350 units) + (Tshs5,000 × 100 units) + (Tshs3,000 x 200 units + (Tshs1,000 × 300 units)

= Tshs800,000 + Tshs100,000 + Tshs2,450,000 + Tshs500,000 + Tshs600,000 + Tshs300,000

= Tshs 4,750,000

4.4 Least Cost Method (LCM) of solving a transportation problem

In this method, the cheapest route is always the focus for allocation. It is a better method compared to NWCR because costs are considered for allocation. The algorithm is stated thus:

The steps to obtain the initial basic feasible solution are as follows:

Step 1 Select the smallest transportation cost cell available in the entire table and allocate the supply and demand.

Step 2 Delete the row/column that is exhausted. The deleted row/column must not be considered for further allocation.

Step 3 Again select the smallest cost cell in the existing table and allocate. (Note: if there are two or more costs that are the least, select the cells where the maximum allocation can be made.)

Step 4 Repeat steps (ii) and (iii) until all the supply and demand values are zero.

Step 5 Obtain the initial basic feasible solution.

Example

The cost of transportation per unit from three sources and four destinations are given below.

Transportation Cost per unit (Tshs'000)

Source	Destination				Supply (units)
	1	2	3	4	
	Tshs'000	Tshs'000	Tshs'000	Tshs'000	
1	4	2	7	3	250
2	3	7	5	8	450
3	9	4	3	1	500
Demand (units)	200	400	300	300	1200

Required:

Obtain the initial basic feasible solution using the Least Cost Method.

Answer

Step 1: Formulate the problem

We need to formulate the problem first and set it up in a matrix form. We need to check whether the problem is a balanced or unbalanced transportation problem.

Total demand		Total supply
$\sum_{j=1}^{n} bi = 1,200$	=	$\sum_{i=1}^{m} ai = 1,200$

The given problem is a balanced one as the demand and supply are equal – i.e. 1200 units.

Step 2: Obtain the initial feasible solution

The initial feasible solution here needs to be obtained by the Least Cost Method.

Step 1: Select the smallest transportation cost cell available in the entire table and allocate the supply and demand.

In the given problem, the least cost cell is (3,4). The corresponding supply and demand values are 500 and 300 respectively. Allocate the maximum possible units.

From the supply value of 500, the demand value of 300 is satisfied. Therefore, the demand at destination 4 is fully satisfied.

The allocation is shown below.

Transportation Cost per unit (Tshs'000)

Source	Destination				Supply (units)
	1	2	3	4	
1	4	2	7	3	250
2	3	7	5	8	450
3	9	4	3	$\boxed{300}$ 1	200
Demand (units)	200	400	300	~~300~~	

Continued on the next page

Step 2: Delete that row/column which has been exhausted. The deleted row/column must not be considered for further allocation.

Here the demand at destination 4 is fully satisfied. Therefore column 4 needs to be deleted. The result is shown in the table given below:

Table after deleting column 4

Transportation Cost per unit (Tshs'000)

Source	Destination			Supply (units)
	1	2	3	
1	4	2	7	250
2	3	7	5	450
3	9	4	3	200
Demand (units)	200	400	300	

Step 3: Again select the smallest cost cell in the existing table and allocate. (Note: if there are two or more costs that are the least, select the cells where the maximum allocation can be made.)

Here, the least cost cell in the table given above is (1,2). The corresponding values of demand and supply are 400 and 250. We need to allocate the demand to the total supply at source 1 which gives us a remaining demand of 150 at destination 2. Now, we need to delete row 1 as shown below:

Transportation Cost per unit (Tshs'000)

Source	Destination			Supply (units)
	1	2	3	
1	4	250 2	7	~~250~~
2	3	7	5	450
3	9	4	3	200
Demand (units)	200	150	300	

Table after deleting Row 1

Transportation Cost per unit (Tshs'000)

Source	Destination			Supply (units)
	1	2	3	
2	3	7	5	450
3	9	4	3	~~500~~ 200
Demand (units)	200	~~400~~ 150	300	

Step 4: Repeat steps (ii) and (iii) until all the supply and demand values are zero.

Now, in the table given above, there are two least cost cells - (2,1) and (3,3). In such a situation, the cell where the maximum allocation can be made should be selected. In this case, the maximum allocation in both cells is 200, so any cell can be selected.

Continued on the next page

Here we select cell (2,1) and the corresponding values of demand are supply are 200 and 450. Hence the demand at destination 1 is completely allocated and exhausted as shown in the table given below:

Transportation Cost per unit (Tshs'000)

Source	Destination			Supply (units)
	1	2	3	
2	200 3	7	5	~~450~~ 250
3	9	4	3	~~52~~00
Demand (units)	~~200~~	150	300	

The demand at destination 1 is completely exhausted so we delete column 1 as shown below:

Table after deleting Column 1

Transportation Cost per unit (Tshs'000)

Source	Destination		Supply (units)
	2	3	
2	7	5	250
3	4	3	200
Demand (units)	150	300	

Now, again we have cell (3,3) as the least cost cell and the corresponding values of demand and supply are 300 and 200. The supply of source 3 is fully allocated to the demand as shown below:

Transportation Cost per unit (Tshs'000)

Source	Destination		Supply (units)
	2	3	
2	7	5	250
3	4	200 3	~~200~~
Demand (units)	150	~~300~~ 100	

Since the supply of source 3 is fully exhausted, we can delete row 3 as shown in the table given below.

Table after deleting Row 3

Transportation Cost per unit (Tshs'000)

Source	Destination		Supply (units)
	2	3	
2	7	5	250
Demand (units)	150	100	

Furthermore, we have remaining demand and supply values of 250 each, so the supply is accordingly allocated and we have demand and supply values as zero as shown below.

Transportation Cost per unit (Tshs'000)

Source	Destination		Supply (units)
	2	3	
2	150 7	100 5	~~250~~
Demand (units)	~~150~~	~~100~~	

Continued on the next page

Step 5: Obtain the initial basic feasible solution.

Transportation Cost

= (Tshs2,000 × 250 units)+ (Tshs3,000 × 200 units) + (Tshs7,000 × 150 units) + (Tshs5,000 × 100 units)+ (Tshs3,000 × 200 units) + (Tshs1,000 × 300 units)

= Tshs500,000 + Tshs600,000 + Tshs1050,000 + Tshs500,000 + Tshs600,000 + Tshs300,000
=Tshs 3,550,000

4.5 Vogel's Approximation Method (VAM) for solving transportation problems

VAM, which is also called the penalty method, is an improvement on the LCM method and generates a better initial solution. It makes use of opportunity cost (penalty) principles in order to make allocation to various cells by minimizing the penalty cost.

The steps to obtain the initial basic feasible solution are as follows:

Step 1 Calculate penalties for each row and column by taking the difference between the smallest cost and the next highest cost available in that row/column. If there are two smallest costs, then the penalty is zero.

Step 2 Select the row/column, which has the largest penalty and make allocation in the cell having the least cost in the selected row/column. If two or more equally small penalties exist, select one where a row/column contains minimum unit cost. If there is a tie, select one where maximum allocation can be made.

Step 3 Delete the row/column which has satisfied the supply and demand.

Step 4 Repeat steps (i) and (ii) until the entire supply and demands are satisfied.

Step 5 Obtain the initial basic feasible solution.

Vogel's Approximation Method

Example

The cost of transportation per unit from three sources and four destinations are given below.

Transportation Cost per unit (Tshs'000)

Source	Destination				Supply (units)
	1	2	3	4	
1	4	2	7	3	250
2	3	7	5	8	450
3	9	4	3	1	500
Demand (units)	200	400	300	300	1200

Required:

Obtain the initial basic feasible solution using Vogel's Approximation Method.

Answer

Step 1: Formulate the problem

We need to formulate the problem first and set it up in a matrix form. We need to check whether the problem is a balanced or unbalanced transportation problem.

Total demand		Total supply
$\sum_{j=1}^{N} b_i = 1200$	=	$\sum_{i=1}^{M} a_i = 1200$

The given problem here is a balanced one as the demand and supply is equal; that is 1200 units.

Continued on the next page

Step 2: Obtain the initial feasible solution

The initial feasible solution here needs to be obtained by Vogel's Approximation Method

Step 1: Calculate penalties for each row and column by taking the difference between the smallest cost and the next highest cost available in that row/column. If there are two smallest costs, then the penalty is zero.

The penalty for row 1 is 1 because the least cost in the row is 2 and the next highest cost is 3; and 3-2= (1). In the same way, the penalty for all the rows and columns is calculated.

Transportation Cost per unit (Tshs'000)

Source	Destination				Supply (units)	Penalty
	1	2	3	4		
1	4	2	7	3	250	(1)
2	3	7	5	8	450	(2)
3	9	4	3	1	500	(2)
Demand (units)	200	400	300	300		
Penalty	(1)	(2)	(2)	(2)		

Step 2: Select the row/column which has the largest penalty and make allocation in the cell having the least cost in the selected row/column. If two or more equal penalties exist, select one where a row/column contains the minimum unit cost. If there is a tie, select one where maximum allocation can be made.

In this case, there are five penalties which have the maximum value 2. The cell with least cost is Row 3 and hence select cell (3,4) for allocation. The supply and demand are 500 and 300 respectively and hence we need to allocate 300 in cell (3,4) as shown in the table.

Transportation Cost per unit (Tshs'000)

Source	Destination				Supply	Penalty
	1	2	3	4		
1	4	2	7	3	250	(1)
2	3	7	5	8	450	(2)
3	9	4	3	$\boxed{300}$ 1	~~500~~ 200	(2)
Demand	200	400	300	~~300~~		
Penalty	(1)	(2)	(2)	(2)		

Continued on the next page

Step 3: Delete the row/column, which has satisfied the supply and demand.

Now, the demand for destination 4 is fully satisfied so we can delete column 4 as shown in below.

Transportation Cost per unit (Tshs'000)

Source	Destination			Supply	Penalty
	1	2	3		
1	4	2	7	250	(1)
2	3	7	5	450	(2)
3	9	4	3	~~500~~ 200	(2)
Demand	200	400	300		
Penalty	(1)	(2)	(2)		

Step 4: Repeat steps (i) and (ii) until the entire supply and demands are satisfied.

Now, the penalties are calculated for the remaining rows and columns as shown below.

Transportation Cost per unit (Tshs'000)

Source	Destination			Supply	Penalty
	1	2	3		
1	4	2	7	250	(2)
2	3	7	5	450	(2)
3	9	4	3	~~500~~ 200	(1)
Demand	200	400	300		
Penalty	(1)	(2)	(2)		

Here, there are four maximum penalty values of 2. We need to select the corresponding least cost cell, (1,2) which has the least unit transportation cost of 2. The cell (1, 2) is selected for allocation and corresponding supply of 250 is allocated to the demand of 400 as shown below:

Transportation Cost per unit (Tshs'000)

Source	Destination			Supply	Penalty
	1	2	3		
1	4	250 2	7	~~250~~	(2)
2	3	7	5	450	(2)
3	9	4	3	~~500~~ 200	(1)
Demand	200	~~400~~ 150	300		
Penalty	(1)	(2)	(2)		

Now, the supply of source 1 is exhausted therefore we need to delete row 1 as shown below.

Continued on the next page

Table after deleting Row 1

Transportation Cost per unit (Tshs'000)

Source	Destination			Supply	Penalty
	1	2	3		
2	3	7	5	450	(2)
3	9	4	3	~~500~~ 200	(1)
Demand	200	~~400~~ 150	300		
Penalty	(1)	(2)	(2)		

Now again we have calculated the penalties as shown below:

Transportation Cost per unit (Tshs'000)

Source	Destination			Supply	Penalty
	1	2	3		
2	3	7	5	450	(2)
3	9	4	3	~~500~~ 200	(1)
Demand	200	~~400~~ 150	300		
Penalty	(6)	(3)	(2)		

Here the maximum penalty value (6) is for column 1 and the least cost cell (2,1) is selected for allocation as shown below:

Transportation Cost per unit (Tshs'000)

Source	Destination			Supply	Penalty
	1	2	3		
2	$\boxed{200}$ 3	7	5	450 250	(2)
3	9	4	3	~~500~~ 200	(1)
Demand	~~200~~	~~400~~ 150	300		
Penalty	(6)	(3)	(2)		

Now, the demand at destination 1 is fully exhausted therefore we need to delete column 1 as shown below:

Transportation Cost per unit (Tshs'000)

Source	Destination		Supply	Penalty
	2	3		
2	7	5	~~450~~ 250	(2)
3	4	3	~~500~~ 200	(1)
Demand	~~400~~ 150	300		
Penalty	(3)	(2)		

Continued on the next page

Again we calculate the penalties as shown in the table below:

Transportation Cost per unit (Tshs'000)

Source	Destination		Supply	Penalty
	2	3		
2	7	5	~~450~~ 250	(2)
3	4	3	~~500~~ 200	(1)
Demand	~~400~~ 150	300		
Penalty	(3)	(2)		

Here the column having the maximum penalty of (3) is Column 2 and the least cost cell selected for allocation is (3,2), while the demand of 150 is fully allocated by the supply as shown below:

Transportation Cost per unit (Tshs'000)

Source	Destination		Supply	Penalty
	2	3		
2	7	5	~~450~~ 250	(2)
3	150 4	3	~~500~~ ~~200~~ 50	(1)
Demand	~~400~~ ~~150~~ 0	300		
Penalty	(3)	(2)		

Since the demand at destination 2 is fully satisfied we can delete column 2 as shown below:

Transportation Cost per unit (Tshs'000)

Source	Destination	Supply	Penalty
	3		
2	5	~~450~~ 250	(2)
3	3	~~500~~ ~~200~~ 50	(1)
Demand	300		
Penalty	(2)		

Now, finally we get only one column, so we cannot calculate penalties. Here the supply of 250 is allocated to the demand and we get zero values for demand and supply.

Transportation Cost per unit (Tshs'000)

Source	Destination	Supply
	3	
2	250 5	~~250~~ 0
3	50 3	~~50~~ 0
Demand	~~300~~	

Continued on the next page

Step 5: Obtain the initial basic feasible solution.

Transportation cost

= (Tshs2,000 × 250 units) + (Tshs3,000 × 200 units) + (Tshs5,000 × 250 units) + (Tshs4,000 × 150 units) + (Tshs3,000 × 50 units) + (Tshs1,000 × 300 units)
= Tshs500,000 + Tshs600,000 + Tshs1250,000 + Tshs600,000+ Tshs150,000 + Tshs300,000
= Tshs 3,400,000

Now let us learn to solve an unbalanced transportation problem with the help of an example:

Example

Given below is a table of demand of a company at its 3 warehouses as destinations and supply of produced units at 2 plants as sources. Also included are the cost per unit incurred to transport one unit from any particular source to any particular destination:

Transportation Cost per unit (Tshs'000)

Source	Destination			Supply
	1	2	3	
1	4	2	7	250
2	3	7	5	450
Demand	200	400	300	

Required:

Calculate the initial basic feasible cost of transporting the units using the Vogel's Approximation Method.

Answer

(All Transportation Costs are per unit costs in Tshs'000)

Step 1: Check whether the problem is a balanced one.

In the given case the problem is an unbalanced one as the demand and supply are as given below:

Total demand		Total supply	Excess
$\sum_{j=1}^{N} b_i = 900$	-	$\sum_{i=1}^{M} a_i = 700$	200 units

The demand exceeds the supply by 200 units.

Therefore we need to insert a dummy row with 200 units and assign zero values to the cost as shown below:

Source	Destination			Supply (units)
	1	2	3	
1	4	2	7	250
2	3	7	5	450
3	0	0	0	200
Demand (units)	200	400	300	

Step 2: Obtain the initial basic feasible solution by the Vogel's Approximation Method

Continued on the next page

Step 2.1: We need to calculate the penalties for the rows and columns as shown below:

Source	Destination			Supply (units)	Penalty
	1	2	3		
1	4	2	7	250	(2)
2	3	7	5	450	(2)
3	0	0	0	200	(0)
Demand (units)	200	400	300		
Penalty	(3)	(2)	(5)		

Step 2.2: In this case, the maximum value of penalty is 5 in Column 3 and the cell with the least cost in Row 3 is (3,3).

The corresponding supply and demand values are 200 and 300 respectively and hence allocate the supply to the demand in cell (3,3) as shown in the table:

Source	Destination			Supply (units)	Penalty
	1	2	3		
1	4	2	7	250	(2)
2	3	7	5	450	(2)
3	0	0	200 0	~~200~~	(0)
Demand (units)	200	400	~~300~~ 100		
Penalty	(3)	(2)	(5)		

Step 2.3: Now, since the entire supply of source 3 is exhausted, we can delete Row 3 as shown below:

Table after deleting Row 3

Source	Destination			Supply (units)	Penalty
	1	2	3		
1	4	2	7	250	(2)
2	3	7	5	450	(2)
Demand (units)	200	400	~~300~~ 100		
Penalty	(3)	(2)	(5)		

Step 2.4: Again we need to calculate the penalties:

Source	Destination			Supply (units)	Penalty
	1	2	3		
1	4	2	7	250	(2)
2	3	7	5	450	(2)
Demand (units)	200	400	~~300~~ 100		
Penalty	(1)	(5)	(2)		

Continued on the next page

Now, here the highest penalty we have is 5 in Column 2 and the least cost cell in the column is (1,2), while the corresponding demand and supply values are 400 and 250. We therefore allocate the supply of source 1 to the demand at destination 2 as shown below:

Source	Destination			Supply (units)	Penalty
	1	2	3		
1	4	250 2	7	~~250~~	(2)
2	3	7	5	450	(2)
Demand (units)	200	~~400~~ 150	~~300~~ 100		
Penalty	(1)	(5)	(2)		

Table after deleting Row 1

Now, since the supply of source 1 is completely exhausted, we need to delete Row 1.

Source	Destination			Supply (units)	Penalty
	1	2	3		
2	3	7	5	450	(2)
Demand (units)	200	~~400~~ 150	~~300~~ 100		
Penalty	(1)	(5)	(2)		

Now only one row is remaining, therefore penalties cannot be calculated. The remaining demand and supply values are 450 each so the supply is allocated to the demands as shown below:

Source	Destination			Supply (units)
	1	2	3	
2	200 3	150 7	100 5	~~450~~
Demand (units)	~~200~~	~~400~~ ~~150~~	~~300~~ ~~100~~	

Total Transportation Cost:

= (200 units x Tshs0) + (250 units x Tshs2,000) + (200 units x Tshs3,000) + (150 units x Tshs7,000) + (100 units x Tshs5,000)

= Tshs0 + Tshs500,000 + Tshs600,000 + Tshs1,050,000 + Tshs500,000 =Tshs2,650,000

4.6 Method for checking optimality and the Stepping Stone Method

This is one of the methods used to determine the optimality of the initial basic feasible solution obtained by the North West Corner Rule, Least Cost Method and the Vogel's Appropriation Method.

If we are given a task to cross a pond, what would our strategy be?

We would use stones lying in the pond to step on them and cross the pond. The same strategy is used by this method. Here the transportation problem is the pond and the cells used are the stepping stones used to cross the pond.

An optimal solution is one in which there are no other transportation routes that would reduce the total transportation cost, for which we have to evaluate each unoccupied cell in the table in terms of opportunity cost. In this process, if there is no negative opportunity cost, the solution is an optimal solution.

The following are the steps to reach an optimal Answer

Step 1: We need to start at an empty/unused/unoccupied cell (the cost cell to which any supply/demand units are not allocated) and trace back to the original cell forming a closed path or loop moving through used/occupied cells (the cost cell to which supply/demand units are allocated).

There are certain conditions to be followed while forming a path:

(a) Only horizontal and vertical moves are allowed, but no diagonal moves are allowed.

(b) Direction can only be changed at used/occupied cells.

(c) No cell appears more than once in a sequence.

Step 2: For every traced path or loop, increase the quantity by multiplying with 1 and then adjust the allocation to satisfy the demand and supply by alternately increasing and decreasing the quantity by multiplying the cost with 1 and -1 at each used cell in the loop.

Step 3: Now, calculate the relative cost coefficient (Cij) by adding all the unit costs in the path and calculate the cost coefficient (Cij) for all the unused cells.

Step 4: If all the relative cost coefficients (Cij) calculated are equal to or greater than zero then an optimal solution has been achieved.

Step 5: If optimal solution has not been achieved then select the loop or path that has the most negative value and use it to further improve the solution. Now, to improve the solution, select the smallest allocated quantity in the cell multiplied with -1 in the selected path. This quantity is added to cells in the closed loop multiplied with 1 and subtracted from cells that are multiplied with -1. This will give a new basic feasible solution.

Step 6: Again repeat steps 1 to 5 until you achieve an optimal solution.

Now let us understand the way to check for optimality for an initial basic feasible solution with the help of an example.

Example

Kawasaki Co. a company manufacturing bikes, has two plants, A and B with a yearly production capacity of 30,000 and 50,000 bikes. It has its showrooms X, Y and W at three different locations with a demand of 20,000, 25,000 and 35,000 bikes during the year.

The cost of carrying the bikes from the plants to the showrooms is as follows:

Transportation Cost per unit (Tshs'000)

Source	Destinations		
	X	Y	W
A	3	5	7
B	4	7	11

Now obtain the initial basic feasible solution by the North West Corner Rule and obtain the optimal solution by the Stepping Stone method.

Answer

	Destination				Supply
		X	Y	W	
Source/Plant	A	20 1	10 3	5	~~30~~ ~~10~~
	B	2	15 5	35 9	~~50~~ 0
Demand		~~20~~ 0	~~25~~ ~~15~~	~~35~~ 0	80/80

Continued on the next page

Initial Basic Feasible Cost

= (20 units x Tshs1,000) + (10 units x Tshs3,000) + (15 units x Tshs5,000) + (35 units x Tshs9,000)

= Tshs20,000 + Tshs30,000 + Tshs75,000 + Tshs315,000

= Tshs440,000

Optimality Test:

Step 1

Closed path or loop at unused cell AW is AW > BW > BY > AY as shown below:

Source/Plant		Destination			Supply (units)
		X	Y	W	
	A	20 1	10 3	5	~~30~~ ~~10~~
	B	2	15 5	35 9	~~50~~ 0
Demand (units)		~~20~~ 0	~~25~~ ~~15~~	~~35~~ 0	80/80

Closed path or loop at unused cell AW is AW > BW > BY > AY

Step 2

AW (1) > BW (-1) > BY (1) > AY(-1)

Step 3

The cost coefficient (Cij)

AW + BW + BY + AY
= 5(1) + 9(-1) + 5(1) + 3(-1)
= 5 - 9 + 5 – 3
= -2

Now Step 1, 2 and 3 for the unused cell BX,

Step 1

Closed path or loop at unused cell BX is BX > AX > AY > BY as shown below:

Source/Plant		Destination			Supply (units)
		X	Y	W	
	A	20 1	10 3	5	~~30~~ ~~10~~
	B	2	15 5	35 9	~~50~~ 0
Demand (units)		~~20~~ 0	~~25~~ ~~15~~	~~35~~ 0	80/80

Step 2

BX (1) > AX (-1) > AY (1) > BY(-1)

Step 3

The cost coefficient (Cij)

= BX (1) + AX (-1) + AY (1) + BY (-1)
= 2 (1) + 1 (-1) + 3 (1) + 5 (-1)
= 2 + (-1) + 3 + (-5)
= 2 – 1 + 3 – 5
= -1

Step 4

Since the cost coefficient cell is less than 0 for both the unused cells, we can say that the optimal solution has not been achieved.

Step 5

The path of cell AW has the most negative value and the smallest quantity allocated in the cell multiplied with -1 is 10.

Therefore, 10 is added to the cells that are multiplied with 1 and subtracted from those cells which are multiplied with -1 as shown in the table below:

		Destination			Supply (units)
		X	Y	W	
Source/Plant	A	20 1	10 - 10 3	5 0 +10	~~30~~ ~~10~~
	B	2	15 +10 5	35 - 10 9	~~50~~ 0
Demand (units)		~~20~~ 0	~~25~~ ~~15~~	~~35~~ 0	80/80

The new allocation matrix will be as shown below:

		Destination			Supply (units)
		X	Y	W	
Source/Plant	A	20 1	3	10 5	~~30~~ ~~10~~
	B	2	25 5	25 9	~~50~~ 0
Demand (units)		~~20~~ 0	~~25~~ ~~15~~	~~35~~ 0	80/80

New basic feasible solution

= (20 x Tshs1,000) + (10 x Tshs5,000) + (25 x Tshs5,000) + (25 x Tshs9,000)
= Tshs20,000 + Tshs50,000 + Tshs125,000 + Tshs225,000
= Tshs420,000

Step 6

We again need to repeat steps 1 to 5 to achieve an optimal solution.

Test Yourself 4

Sony International has 3 plants or locations (A, B, C) where its music systems can be produced with a production capacity of 50, 60, and 40 per month respectively. These units are to be distributed to 4 points of consumption (X, Y, W, and Z) with the demand of 50, 60, 30 and 10 per month respectively.

The following table gives the transportation cost from various plants to the various points of consumption:

Transportation Cost per unit (Tshs'000)

		Destination				Supply (units)
		X	Y	W	Z	
Source/Plant	A	21	18	27	22	50
	B	19	18	24	20	60
	C	24	25	28	25	40
Demand (units)		50	60	30	10	150/150

Required:

Obtain the Initial basic feasible solution by the following methods:

(a) North West Corner Rule

(b) Vogel's Approximation Method

Test Yourself 5

Which of the following is not a method of obtaining the initial basic feasible solution of a transportation problem?

A North West Corner Rule
B Least Cost Method
C Stepping Stone Method
D Simplex Method

5. Solve an assignment problem by allocation that will produce optimal solutions.
Apply the concept of assignment problem to accounting and business situations.
[Learning Outcome g and h]

5.1 Meaning of assignment problem

Definition

An assignment problem is defined as the problem to assign a number of resources to a number of activities so as to minimise the total cost or to maximize the total profit of allocation in such a way that the measure of effectiveness is optimized.

Assignment arises because available resources such as men, machines, etc. have varying degrees of efficiency for performing different activities. Therefore cost, profit or time for performing the different activities is different. Hence, the problem is, how should the assignments be made so as to optimize (maximize or minimise) the given objective. The assignment model can be applied in many decision-making processes like determining optimum processing time in machine operators and jobs, effectiveness of teachers and subjects, designing of good plant layout, etc. This technique is found suitable for working out routes for travelling salesmen to minimise the total travelling cost, or to maximize the sales. For example in a factory there are 4 different types of machines and there are 5 different operators.

5.2 Application of Assignment Models in business activities

(a) Assignment of employees to machines.

(b) Assignment of operators to jobs.

(c) Effectiveness of teachers and subjects.

(d) Allocation of machines for optimum utilization of space.

(e) Allocation of salesmen to different sales areas.

(f) Allocation of clerks to various counters.

5.3 Mathematical formulation and developing Linear programming Model for an assignment problem

An assignment problem is a special class of linear programming problems. The objective of an assignment problem is to determine the optimal assignment of given tasks to a set of workers who can perform the tasks with varying efficiency, in terms of time taken, cost, amount of sales etc.

Consider a business unit that has some machines which need to be operated manually; and it also has operators which can operate these machines. These operators have varying degrees of efficiency in performing the job. Therefore cost, profit or time for performing the different activities is different. The concern of the business unit is to get the work done by assigning the operators to the machines at the lowest possible cost. This can be done by formulating a linear programming model in the following way:

Let there be n number of jobs and m number of operators and the ith job is assigned to the jth operator.

Now, let C_{ij} be the cost of performing (or the time taken by the operator, which will lead us to cost) the i^{th} job by j^{th} operator and X_{ij} represent the assignment of j^{th} operator to i^{th} job. Now, the problem is to determine the most effective assignment schedule which minimises the total cost termed as Z.

Z can be achieved by adding the cost (Cij) for every job assigned (Xij) to every operator.

Now the above assignment problem can be stated mathematically as a linear programming model below:

$$z = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} c_{ij}$$

Constraints in the Assignment Required

Now, in the given situation there are certain constraints also. One operator can only perform one job at a time. Therefore, the above formula for total cost is subject to formula jobs constraints also.

The formulae for supply and demand constraints are given below:

Operator constraints

$$\sum_{j=1}^{n} x_{ij} = 1$$

where, j = 1,2,3,……..m

Job Constraints

$$\sum_{i=1}^{n} x_{ij} = 1$$

where, I = 1,2,3, ……n

and $X_{ij} \geq 0$ for all i and j.

Now, let us see how to develop a Linear Programming Model for an Assignment Problem with the help of an example:

Example

Given below is the time taken by each operator (in minutes) to perform a job:

Operator	Job		
	1	2	3
	Minutes	Minutes	Minutes
A	10	16	7
B	9	17	6
C	6	13	5

We need to formulate a Linear Programming Model for the above Assignment Problem.

Now Let:

X_{11} represent the assignment of operator A to job 1
X_{12} represent the assignment of operator A to job 2
X_{13} represent the assignment of operator A to job 3
X_{21} represent the assignment of operator B to job 1
And so on ……

Equations for the time taken by each operator:

$10 \times X_{11} + 16 \times X_{12} + 7 \times X_{13}$ = time taken by operator A.
$9 \times X_{21} + 17 \times X_{22} + 6 \times X_{23}$ = time taken by operator B.
$6 \times X_{31} + 13 \times X_{32} + 5 \times X_{33}$ = time taken by operator C.

The constraint in this assignment problem is that each operator must be assigned to only one job and similarly, each job must be performed by only one operator. Taking this constraint into account, the constraint equations are as follows:

$X_{11} + X_{12} + X_{13} = 1$ operator A
$X_{21} + X_{22} + X_{23} = 1$ operator B
$X_{31} + X_{32} + X_{33} = 1$ operator C
$X_{11} + X_{21} + X_{31} = 1$ Job
$X_{12} + X_{22} + X_{32} = 1$ Job
$X_{13} + X_{23} + X_{33} = 1$ Job

Now, the linear programming model for the problem will be:

Minimise $Z = 10 (X_{11}) + 16 (X_{12}) + 7 (X_{13}) + 9 (X_{21}) + 17 (X_{22}) + 6 (X_{23}) + 6 (X_{31}) + 13 (X_{32}) + 5 (X_{33})$

Subject to constraints:

$X_{11} + X_{12} + X_{13} = 1$	(1)
$X_{21} + X_{22} + X_{23} = 1$	(2)
$X_{31} + X_{32} + X_{33} = 1$	(3)
$X_{11} + X_{21} + X_{31} = 1$	(4)
$X_{12} + X_{22} + X_{32} = 1$	(5)
$X_{13} + X_{23} + X_{33} = 1$	(6)

where, $X_{ij} > 0$ for i = 1,2,3 and j = 1,2,3.

Test Yourself 6

The assignment model is a special type of the ________ model.

A Hypothesis
B Transportation
C Integration
D Inventory

5.4 Understanding balanced assignment problems and unbalanced assignment problems

Assignment Problems like Transportation Problems are of two types:

(a) Balanced Assignment Problems
(b) Unbalanced Assignment Problems

In any given real life situation, the number of workers working in a factory are very rarely equal to the number of jobs to be done. Both these elements are affected by many reasons like the workers may be on leave, the factory is in the process of recruitment so the workers are short, the particular skill required for the job is available only in few workers etc. On the other hand, there may be other problems with the job e.g. there might be extra work due to seasonal demand, the work may be urgent due to client request, a machine may break down, there may be slack due to less demand etc. Hence, the numbers of workers required to perform jobs may not always be equal to the jobs to be done. Therefore, we need to understand the concept of balanced and unbalanced assignment problems.

1. Balanced Assignment Problems

If the number of rows is equal to the number of columns or if the given problem is a square matrix, the problem is termed as a balanced assignment problem.

2. Unbalanced Assignment Problems

If the given problem is not a square matrix, the problem is termed as an unbalanced assignment problem.

The number of jobs to be done and the number of operators available are rarely equal in a given practical problem. There are many factors affecting the jobs to be performed. For instance, there may be days when there is a more-than-usual rush at a bank, there may be more work at a factory during peak season, etc. At the same time, there are a number of reasons affecting the availability of operators, like labour strike, some operators being on leave, attrition etc.

5.5 Procedure to convert an Unbalanced Assignment Problem into a Balanced Assignment Problem

Unbalanced Assignment Problems can be easily solved by introducing dummy sources (OPERATORS) and dummy destinations (JOBS) in the following manner:

Step 1: If the total number of jobs are greater than the total number of operators, a dummy operator (dummy row) with operators equal to the job surplus is added.

Step 2: If the total number of operators are greater than the total number of jobs, a dummy job (dummy column) with jobs equal to the operator surplus is added.

Step 3: The time taken for the dummy column and dummy row are assigned zero values, because no job or operator has actually increased.

5.6 Procedure to solve an Assignment Problem

The method usually used to solve an assignment problem is the Hungarian Method, also known as The Reduced Matrix Method.

Steps to be followed to solve an Assignment Problem using the Hungarian Method:

Step 1	In a given problem, check whether the problem is a balanced one or an unbalanced one. If it is an unbalanced one then add a dummy row or a dummy column. The assignment costs for dummy cells are always assigned a zero value.
Step 2	Reduce the values in the total cost table by selecting the smallest value in each row and subtracting it from other values in the corresponding row.
Step 3	Reduce the values in the new total cost table by selecting the smallest value in each column and subtracting it from other values in that column.
Step 4	Draw straight lines vertically or horizontally through the total cost table so that all the zeros in the table are connected using the minimum number of lines possible.
Step 5	If number of lines drawn = order of matrix, then optimality is reached, so proceed to step 7. If optimality is not reached, then go to step 6.
Step 6	(a) Select the smallest number in the table that is not covered by a straight line and subtract this number from all other numbers not covered by a straight line. (b) Add the same lowest number (selected above) to the number lying in the intersection of any two lines. (c) Check whether optimality has been reached by repeating steps 4 and 5.
Step 7	Assign the tasks to the operators. Take any row or column which has a single zero and assign by marking it with a square. Strike off the remaining zeros, if any, in that row and column (X). Repeat the process until all the assignments have been made.

While assigning, if there exists no single zero in the row or column, choose any one zero and assign it. Strike off the remaining zeros in that column or row, and repeat the same for other assignments also. If there is no single zero allocation, it means multiple numbers of solutions exist. But the cost will remain the same for the different sets of allocations.

Let us understand how to solve an assignment problem with the help of an example.

Example

A factory has four machines and four operators. The assigning costs of the operators to the machines are given in the table below:

Tasks	Operators			
	1	2	3	4
	Tshs'000	Tshs'000	Tshs'000	Tshs'000
A	18	26	17	11
B	13	28	29	26
C	38	19	18	15
D	19	26	24	10

Assign the four tasks to the four operators, such that the assigning cost is the minimum.

Answer

Let us see the method of solving this problem, step wise.

Step 1

In a given problem, check whether the problem is a balanced one or an unbalanced one. If it is an unbalanced one then add a dummy row or a dummy column. The assignment costs for dummy cells are always assigned a zero value.

The given problem is a balanced one therefore there is no need to add a dummy row or column.

Continued on the next page

Step 2

Reduce the values in the total cost table by selecting the smallest value in each row and subtracting it from other values in the corresponding row.

In the given problem the smallest value in Row A is 11, Row B is 13, Row C is 15 and Row D is 10. Therefore, the reduced values are as given in the table below:

Tasks	Operators			
	1	2	3	4
A	7	15	6	0
B	0	15	16	13
C	23	4	3	0
D	9	16	14	0

Step 3

Reduce the values in the new total cost table by selecting the smallest value in each column and subtracting it from other values in that column.

In the new matrix, the smallest value in Column 1 is 0, Column 2 is 4, column 3 is 3 and column 4 is 0. Therefore, the reduced values are as given in the table below:

Tasks	Operators			
	1	2	3	4
A	7	11	3	0
B	0	11	13	13
C	23	0	0	0
D	9	12	11	0

Step 4

Draw straight lines vertically or horizontally through the total cost table so that all the zeros in the table are connected by using the minimum number of lines possible.

The matrix with lines connecting the zeros is given in the table below:

Tasks	Operators			
	1	2	3	4
A	7	11	3	0
B	0	11	13	13
C	23	0	0	0
D	9	12	11	0

Step 5

If the number of lines drawn is equal to the number of rows or columns, then optimality is reached, so proceed to step 7. But, if optimality is not reached, we modify the total cost table in step 6.

Here, the number of lines drawn is 3, but the number of rows or columns is 4, therefore optimality has not been reached.

Continued on the next page

Step 6

(a) Select the smallest number in the table that is not covered by a straight line and subtract this number from all other numbers not covered by a straight line.

(b) Add the same lowest number (selected above) to the number lying at the intersection of any two lines.

(c) Check whether optimality has been reached by repeating steps 4 and 5.

Here, the smallest number in the table that is not connected is 3. After subtracting this number from all numbers not covered by any straight line and adding it to the numbers lying at the intersection, the new matrix is as given below:

Tasks	Operators			
	1	2	3	4
A	7	8	0	0
B	0	8	10	13
C	26	0	0	3
D	9	9	8	0

The Total Cost Table after drawing lines is given below:

Tasks	Operators			
	1	2	3	4
A	7	8	0	0
B	0	8	10	13
C	26	0	0	3
D	9	9	8	0

Step 7: Assign the tasks to the operators.

Take any row or column which has a single zero and assign by marking it with a square. Strike off the remaining zeros, if any, in that row and column (X). Repeat the process until all the assignments have been made.

Tasks	Operators			
	1	2	3	4
A	7	8	$\boxed{0}$	~~0~~
B	$\boxed{0}$	8	10	13
C	26	$\boxed{0}$	~~0~~	3
D	9	9	8	$\boxed{0}$

Therefore, the optimal assignment is as follows:

Task	Operator	Cost
		Tshs'000
A	3	17
B	1	13
C	2	19
D	4	10
Total Cost		59

Test Yourself 7

In a textile manufacturing unit, there five machines that are used for four jobs. Each job can be assigned to one and only one machine. The cost of each job on each machine is given in the following table:

Job	Machines				
	A	B	C	D	E
	Tshs'000	Tshs'000	Tshs'000	Tshs'000	Tshs'000
1	7	9	13	8	9
2	10	7	7	8	7
3	8	9	12	9	5
4	12	6	10	4	6

Required:

Assign the jobs to the machines such that the cost of performing the tasks is the minimum.

Test Yourself 8

Which of the following is a method of solving an Assignment Problem?

A North West Corner Rule
B Least Cost Method
C Stepping Stone Method
D Hungarian Method

Answers to Test Yourself

Answer to TY 1

The correct option is C.

The main objective of a transportation problem is to distribute a single good from various sources to different destinations at minimum total cost

Answer to TY 2

The correct option is C.

Answer to TY 3

The correct option is A.

Answer to TY 4

Initial Basic Feasible solution by North West Corner Rule:

		Destination				Supply (units)
		X	Y	W	Z	
Source/Plant	A	$\boxed{50}$ 21	18	27	22	~~50~~ 0
	B	19	$\boxed{60}$ 18	24	20	~~60~~ 0
	C	24	25	$\boxed{30}$ 28	$\boxed{10}$ 25	~~40~~ 0
Demand (units)		~~50~~ 0	~~60~~ 0	~~30~~ 0	~~10~~ 0	150/150

Total Cost

= (50 units x Tshs21,000) + (60 units x Tshs18,000) + (30 units x Tshs28,000) + (10 units x Tshs25,000)

= Tshs1,050,000 + Tshs1,080,000 + Tshs840,000 + Tshs250,000

= Tshs3220,000

Initial Basic Feasible solution by Vogel's approximation Method:

Calculation of penalty and allocation of demand and supply for the first time:

Source/Plant	Destination					Supply (units)	Penalty
		X	Y	W	Z		
	A	21	$\boxed{50}$ 18	27	22	~~50~~ 0	(3)
	B	19	18	24	20	60	(1)
	C	24	25	28	25	40	(1)
Demand (units)		50	~~60~~ 10	30	10	150/150	
Penalty		(2)	(0)	(3)	(2)		

Calculation of penalty and allocation of demand and supply after deletion of Row A:

Source	Destination					Supply (units)	Penalty
		X	Y	W	Z		
	B	19	$\boxed{10}$ 18	24	20	~~60~~ 50	(1)
	C	24	25	28	25	40	(1)
Demand (units)		50	~~60~~ ~~10~~	30	10	150/150	
Penalty		(5)	(7)	(4)	(5)		

Calculation of penalty and allocation of demand and supply after deletion of Column Y:

Source/Plant	Destination				Supply (units)	Penalty
		X	W	Z		
	B	$\boxed{50}$ 19	24	20	~~60~~ ~~50~~	(1)
	C	24	28	25	40	(1)
Demand (units)		~~50~~	30	10	150/150	
Penalty		(5)	(4)	(5)		

Calculation of penalty and allocation of demand and supply after deletion of Row B and Column X:

Source/ Plant		Destination		Supply (units)
		W	Z	
	B	24	20	~~60~~ ~~50~~ 0
	C	30 28	10 25	~~40~~ 0
Demand (units)		~~30~~ 0	~~10~~ 0	150/150

Total Cost

= (50 units x Tshs18,000) + (50 units x Tshs19,000) + (10 units x Tshs18,000) + (30 units x Tshs28,000) + (10 units x Tshs 25,000)

= Tshs900,000 + Tshs950,000 + Tshs180,000 + Tshs840,000 + Tshs250,000

= Tshs3,120,000

Answer to TY 5

The correct option is D.

Answer to TY 6

The correct answer is B.

The assignment model is a special type of the transportation model.

Answer to TY 7

Step 1

In a given problem, check whether the problem is a balanced one or an unbalanced one. If it is an unbalanced one then add a dummy row or a dummy column. The assignment costs for dummy cells are always assigned as zero.

The given problem is an unbalanced one. The number of machines is less than the number of jobs. Therefore we need to insert a dummy row, D5 as shown below:

Job	Machines				
	A	B	C	D	E
	Tshs'000	Tshs'000	Tshs'000	Tshs'000	Tshs'000
1	7	9	13	8	9
2	10	7	7	8	7
3	8	9	12	9	5
4	12	6	10	4	6
D5	0	0	0	0	0

Step 2

Reduce the values in the total cost table by selecting the smallest value in each row and subtracting it from other values in the corresponding row.

In the given problem, the smallest value in Row 1 is 7, Row 2 is 7, Row 3 is 5, Row 4 is 4 and Row D5 is 0. Therefore, the reduced values are as given in the table below:

Job	Machines				
	A	B	C	D	E
	Tshs'000	Tshs'000	Tshs'000	Tshs'000	Tshs'000
1	0	2	6	1	2
2	3	0	0	1	0
3	3	4	7	4	0
4	8	2	6	0	2
D5	0	0	0	0	0

Step 3

Reduce the values in the new total cost table by selecting the smallest value in each column and subtracting it from other values in that column.

In the new matrix, the smallest value in Column A is 0, Column B is 0, column C is 0, column D is 0 and column E is 0.

Therefore, the reduced values are as given in the table below:

Job	Machines				
	A	B	C	D	E
	Tshs'000	Tshs'000	Tshs'000	Tshs'000	Tshs'000
1	0	2	6	1	2
2	3	0	0	1	0
3	3	4	7	4	0
4	8	2	6	0	2
D5	0	0	0	0	0

Step 4

Draw straight lines vertically or horizontally through the total opportunity cost in such a manner as to minimise the number of lines necessary to cover all zero cells.

Job	Machines				
	A	B	C	D	E
	Tshs'000	Tshs'000	Tshs'000	Tshs'000	Tshs'000
1	0	2	6	1	2
2	3	0	0	1	0
3	3	4	7	4	0
4	8	2	6	0	2
D5	0	0	0	0	0

Step 5

If the number of lines drawn is equal to the number of rows or columns, then optimality is reached, so proceed to step 7. But, if optimality is not reached, or if an optimal assignment is not feasible, we modify the total cost table in step 6.

Here, the number of lines drawn is 5 which are equal to the number of rows or columns i.e. 5, therefore optimality has been reached.

Step 7: Assign the tasks to the operators.

Take any row or column which has a single zero and assign by marking it with a square. Strike off the remaining zeros, if any, in that row and column (X). Repeat the process until all the assignments have been made.

Job	Machines				
	A	B	C	D	E
	Tshs'000	Tshs'000	Tshs'000	Tshs'000	Tshs'000
1	$\boxed{0}$	2	6	1	2
2	3	0	$\boxed{0}$	1	0
3	3	4	7	4	$\boxed{0}$
4	8	2	6	$\boxed{0}$	2
D5	0	0	0	0	0

Note: In this problem there is no zero in Row 4; therefore, there exists multiple solutions to this problem. However, all the solutions shall have the same cost.

Hence, the optimal solution is:

Job	Machine	Cost
		Tshs'000
1	A	7
2	C	7
3	E	5
4	D	4
5	C	0
Total Cost		23

This is the optimal solution to this problem.

Answer to TY 8

The correct option is D

The Hungarian Method is an appropriate method to solve assignment problems.

Self Examination Questions

Question 1

A state has 3 power plants with generating capacities of 50, 60, and 40 million KWH. These plants supply power to four different cities. The demand requirement of the cities is 50, 60, 30 and 10 million KWH respectively.

The following table gives the distribution cost per million unit of power from the three different cities:

Distribution Cost per million units (Tshs'000)

	Destination					Supply (units)
		X	Y	W	Z	
Source/Plant	A	21	18	27	22	50
	B	19	18	24	20	60
	C	24	25	28	25	40
Demand (units)		50	60	30	10	150/150

Determine the initial basic feasible solution for distributing power by the Least Cost Method.

A Tshs3,120,000
B Tshs3,130,000
C Tshs3,320,000
D Tshs3,180,000

Question 2

A company has plant locations A and B with daily capacities of production of 1,000 kg and 500 kg respectively. The cost of production (per kg) is Tshs8,000 and Tshs7,500 respectively. The company has three customers whose requirements of chemicals per day are 400 kg, 700 kg and 500 kg respectively. Transportation cost per kg from plant locations to the customers' premises is given below:

Transportation Cost per kg (Tshs'000)

		Customer		
Plant		1	2	3
	A	5	7	10
	B	4	6	3

Required:

(a) Determine whether the problem is:

A A balanced transportation problem
B An unbalanced transportation problem

(b) Determine the initial basic feasible solution for transporting the chemicals at the least cost possible using the North West Corner Rule.

A Tshs3,700,000
B Tshs8,000,000
C Tshs7,600,000
D Tshs7,700,000

Question 3

Whirfool Inc manufactures refrigerators at 2 different locations in Tanzania in Dares salaam and Tanga with production capacities of 100 and 150 units per month. It has its distributors at 3 different locations in Mbeya, Tabora and Singida with demand for 80, 100 and 70 units per month respectively. The cost of transporting these refrigerators to the distributors is given below:

Transportation cost per unit in Tshs ('000)

		Distributors		
		Mbeya	Tabora	Singida
Plants	Dares salaam	65	45	25
	Tanga	55	35	20

Required:

Determine an optimal transportation schedule and cost for Whirlfool Inc.

Question 4

A company has five machines that are used for four jobs. Each job can be assigned to one and only one machine. The cost of each job on each machine is given in the following table:

Cost in Tshs ('000)

		Machines				
		I	II	III	IV	V
Job	A	4	6	10	5	6
	B	7	4	4	5	4
	C	5	6	9	6	2
	D	9	3	7	1	3

Required:

Assign the machines to the jobs in the most cost effective manner, and determine the optimal total cost.

A Tshs10,000
B Tshs11,000
C Tshs12,000
D Tshs13,000

Question 5

The following matrix depicts the processing time required by each worker for the respective jobs in terms of hours.

	Jobs			
Men	I	II	III	IV
X	20	15	25	10
Y	22	24	26	28
W	18	21	22	25
Z	15	18	20	15

Required:

Solve the assignment problem in the above table using the Hungarian method.

Answers to Self Examination Questions

Answer to SEQ 1

The correct option is A.

Initial Basic Feasible solution by Least Cost Method:

Distribution Cost per million units (Tshs'000)

		Destination				Supply
		X	Y	W	Z	
Source/Plant	A	21	$\boxed{50}$ 18	27	22	~~50~~ 0
	B	$\boxed{50}$ 19	$\boxed{10}$ 18	24	20	~~60~~ ~~50~~ 0
	C	24	25	$\boxed{30}$ 28	$\boxed{10}$ 25	~~40~~
Demand		~~50~~	~~60~~ ~~10~~ 0	~~30~~	~~10~~	150/150

Total Cost

= (50 units x Tshs18,000) + (50 units x Tshs19,000) + (10 units x Tshs18,000) + (30 units x Tshs28,000) + (10 units + Tshs25,000)

= Tshs900,000 + Tshs950,000 + Tshs180,000 + Tshs840,000 + Tshs250,000

= Tshs 3,120,000

Answer to SEQ 2

(a) The correct option is B.

Transportation Cost per kg (Tshs'000)

		Customer			Supply (in kg)
Plant		1	2	3	
	A	5	7	10	1,000
	B	4	6	3	500
Demand (in kg)		400	700	500	1,500 / 1,600

In the given problem, the supply quantity is 1,500 kg but the demand quantity is 1,600 kg. Therefore, this problem is an unbalanced problem.

To convert the unbalanced problem into a balanced one, a dummy row is added, as shown in the table below:

(b) The correct option is B.

Transportation Cost per kg (Tshs'000)

		Customer			Supply (in kg)
Plant		1	2	3	
	A	5	7	10	1,000
	B	4	6	3	500
	D	0	0	0	100
Demand (in kg)		400	700	500	1,600 / 1,600

Initial basic solution by the North West Corner Rule:

Transportation Cost per kg (Tshs'000)

		Customer			Supply (in kg)
Plant		1	2	3	
	A	$\boxed{400}$ 5	$\boxed{600}$ 7	10	~~1000~~ ~~600~~
	B	4	$\boxed{100}$ 6	$\boxed{400}$ 3	~~500~~ ~~400~~
	D	0	0	$\boxed{100}$ 0	~~100~~
Demand (in kg)		~~400~~	~~700~~ ~~100~~	~~500~~	1,600 / 1,600

Total Transportation cost

= (400 kgs x Tshs5,000) + (600 kgs x Tshs7,000) + (100 kgs x Tshs6,000) + (400 kgs x Tshs3,000) + (100 kgs x Tshs0)

= Tshs2,000,000 + Tshs4,200,000 + Tshs600,000 + Tshs1,200,000 + Tshs0

= Tshs 8,000,000

Answer to SEQ 3

Transportation cost per unit in Tshs ('000)

		Distributors			Supply (units)
		Mbeya	Tabora	Singida	
Plants	Dares salaam	65	45	25	100
	Tanga	55	35	20	150
	Demand (units)	80	100	70	

Initial calculation of penalties and the allocation of demand and supply:

Transportation cost per unit in Tshs ('000)

		Distributors			Penalties	Supply (units)
		Mbeya	Tabora	Singida		
Plants	Dares salaam	65	45	$\boxed{70}$ 25	(20)	~~100~~ 30
	Tanga	55	35	20	(15)	150
	Penalties	(10)	(10)	(5)		
	Demand (units)	80	100	~~70~~		

Calculation of penalties and allocation of demand and supply after deleting the column 'Singida'

Transportation cost per unit in Tshs ('000)

		Distributors		Penalties	Supply (units)
		Mbeya	Tabora		
Plants	Dares salaam	65	$\boxed{30}$ 45	(20)	~~100 30~~
	Tanga	$\boxed{80}$ 55	$\boxed{70}$ 35	(20)	~~150 70 80~~
	Penalties	(10)	(10)		
	Demand (units)	~~80~~	~~100~~		

Total Transportation cost:

= (70 units x Tshs25,000) + (80 units x Tshs55,000) + (30 units x Tshs45,000) + (70 units x Tshs35,000) +

= Tshs1,750,000 + Tshs4,400,000 + Tshs1,350,000 + Tshs2,450,000

= Tshs 9,950,000

Optimality test

		Distributors			Supply (units)
		Mbeya (A)	Tabora (B)	Singida (C)	
Plants	Dares salaam (1)	65	$\boxed{30}$ 45	$\boxed{70}$ 25	~~100~~
	Tanga (2)	$\boxed{80}$ 55	$\boxed{70}$ 35	20	~~150 80~~
	Demand (units)	~~80~~	~~100 20~~		

In the given solution, there are two unused cells, 2A and 2C.

Closed path or loop at unused cell 2A is 2A > 1A > 1B > 2B

Now by multiplying the cells by 1 and -1 we get,

2A (1) > 1A(-1) > 1B(1) > 2B(-1)

Therefore, the cost coefficient

= 2A(1) + 1A(-1) + 1B(1) + 2B(-1)
= 55(1) + 65(-1) + 45(1) + 35(-1)
= 55 + (-65) + 45 + (-35)
= 55 – 65 + 45 – 35
= 0

Closed path or loop at unused cell 2C is 2C > 1C > 1B > 2B

Now by multiplying the cells by 1 and -1 we get,

2C (1) > 1C(-1) > 1B(1) > 2B(-1)

Therefore, the cost coefficient

= 2C(1) + 1C(-1) + 1B(1) + 2B(-1)
= 20(1) + 25(-1) + 45(1) + 35(-1)
= 20 + (-25) + 45 + (-35)
= 20 – 25 + 45 – 35
= 5

Since the cost coefficient for both the cells is 0 and 5 i.e. greater than 0, the initial basic feasible solution is an optimal solution.

Answer to SEQ 4

The given problem is an unbalanced one. Therefore, we need to convert it into a square matrix by adding a dummy row as shown below:

Cost in Tshs ('000)

		Machines				
		I	II	III	IV	V
Job	A	4	6	10	5	6
	B	7	4	4	5	4
	C	5	6	9	6	2
	D	9	3	7	1	3
	E	0	0	0	0	0

Row-wise deduction table:

Cost in Tshs ('000)

		Machines				
		I	II	III	IV	V
Jobs	A	0	2	6	1	2
	B	3	0	0	1	0
	C	3	4	7	4	0
	D	8	2	6	0	2
	E	0	0	0	0	0

In this problem, column-wise deduction is not necessary as every column already has a zero value.

Lines drawn to connect zeros:

Cost in Tshs ('000)

		Machines				
		I	II	III	IV	V
Job	A	0	2	6	1	2
	B	3	0	0	1	0
	C	3	4	7	4	0
	D	8	2	6	0	2
	E	0	0	0	0	0

Number of lines drawn = Order of matrix. Hence the solution is optimal.

The jobs shall be assigned to the machines as shown in the table below:

Cost in Tshs ('000)

		Machines				
		I	II	III	IV	V
Job	A	$\boxed{0}$	2	6	1	2
	B	3	0	$\boxed{0}$	1	0
	C	3	4	7	4	$\boxed{0}$
	D	8	2	6	$\boxed{0}$	2
	E	0	$\boxed{0}$	0	0	0

The optimal solution and the cost is as given below

Job	Machine	Cost
		Tshs
A	I	4,000
B	III	4,000
C	V	2,000
D	IV	1,000
E	II	NIL
	Total	11,000

Note: There can be more than one ways of assigning machines to the jobs in this problem, but that will not affect the total cost. In whichever way we make the assignments, the optimal cost shall remain the same.

Answer to SEQ 5

Row-wise deductions are shown below:

Cost in Tshs ('000)

	Jobs			
Men	I	II	III	IV
X	10	5	15	0
Y	0	2	4	6
W	0	3	4	7
Z	0	3	5	0

Column-wise deductions are shown in the table below:

Cost in Tshs ('000)

	Jobs			
Men	I	II	III	IV
X	10	3	11	0
Y	0	0	0	6
W	0	1	0	7
Z	0	1	1	0

Lines drawn to join the zeros are shown below:

Cost in Tshs ('000)

	Jobs			
Men	I	II	III	IV
X	10	3	11	0
Y	0	0	0	6
W	0	1	0	7
Z	0	1	1	0

Here,

Number of lines drawn = the order of the matrix

Therefore, this is an optimal solution.

Assignment of Men to Jobs is shown below:

Cost in Tshs ('000)

	Job			
Men	I	II	III	IV
X	10	3	11	$\boxed{0}$
Y	0	$\boxed{0}$	0	6
W	0	1	$\boxed{0}$	7
Z	$\boxed{0}$	1	1	0

Total cost

Cost in Tshs ('000)

Men	Job	Cost
X	IV	10
Y	II	24
W	III	22
Z	I	15
Total		71

NETWORK ANALYSIS AND PROJECT SCHEDULING

10

Study Guide 10 -A: ACTIVITY, NODE, CPM AND PERT CONSTRUCTION OF NETWORK

Get Through Intro

Network analysis is a vital technique in PROJECT MANAGEMENT. It enables taking a systematic quantitative structured approach to the problem of managing a project through to successful completion. It has a graphical representation that means it can be understood and used by those with a less technical background.

A Schedule Network Analysis is a graphical representation of a schedule showing each sequenced activity and the time it takes to finish each one. Networks model the interrelated flows of work that must be accomplished to complete a project. They visually portray the events and activities that are planned for the project and show their sequential relationships and interdependencies. Networks generally flow from left to right and may or may not be drawn to scale on a time-based calendar.

Two different techniques for network analysis were developed independently in the late 1950's - these were:

a) PERT (for Program Evaluation and Review Technique); and
b) CPM (for Critical Path Management).

CPM/PERT or Network Analysis as the technique is sometimes called, developed along two parallel streams, one industrial and the other military. PERT was developed in 1958 to aid the US Navy in the planning and control of its Polaris missile program. This was a project to build a strategic weapons system, namely the first submarine launched intercontinental ballistic missile, at the time of the Cold War between the USA and Russia. The DuPont Company and Remington Rand Univac developed CPM in the 1950 as a result of a joint effort. As these were commercial companies cost was an issue, unlike the Polaris project mentioned above. In CPM the emphasis is on the trade-off between the cost of the project and its overall completion time (e.g. for certain activities it may be possible to decrease their completion times by spending more money - how does this affect the overall completion time of the project?)

Learning Outcomes

a) Define what is a project
b) Define Activity
c) Manage to draw a network diagram for both using CPM and PERT.
d) Interpret the Network diagram.

a) Definitions

a) A project
One definition of a project (from the Project Management Institute) is that, a project is a temporary

endeavour undertaken to create a "unique" product or service. This definition highlights two most essential features of a project:

i) it is temporary - A project has a definite beginning and end.
ii) it is "unique" in some way. Every project should have a well-defined objective. This means a Project in "non-repetitive" or "non-routine", e.g. building the very first Boeing Jumbo jet was a project - building them now is a repetitive/routine manufacturing process, not a project.

Projects can be large or small and involve one person or thousands of people. They can be done in one day or take years to complete. Typically, all projects can be broken down into:
i) separate activities (tasks/jobs) - where each activity has an associated duration or completion time (i.e. the time from the start of the activity to its finish)
ii) precedence relationships - which govern the order in which we may perform the activities, e.g. in a project concerned with building a house the activity "erect all four walls" must be finished before the activity "put roof on" can start and the problem is to bring all these activities together in a coherent fashion to complete the project.

Examples of Projects

You can think of many projects in real-life such as:
1. Building the Kigamboni Julius Kambarage Nyerere Bridge
2. Building or Expansion of the Mwalimu Nyerere International airport in Dar-es-Salaam.
3. Building a new Mwongozo Ship in Lake Victoria
4. A team of students creates a smartphone application and sells it online.
5. A global bank acquires a smaller financial institution and needs to reconcile systems and procedures into a common entity.
6. A company develops a new system to increase sales force productivity and customer relationship management that will work on various laptops, smartphones, and tablets.

b) Define An Activity

An activity is a scheduled phase in a project plan with a distinct beginning and end. An activity usually contains several tasks upon completion of which the whole activity is completed. Several activities can be combined to form a summary activity. The duration of an activity is determined by the effort it takes to complete each of its designated tasks.

Activity definition is the process of identifying the specific activities that the project team members and stakeholders must perform to produce the project deliverables. The start or end of an activity can also be determined by constraints, so called dependencies, between activities. It should also be noted that apart from duration /time Activities may need other types of recourses such as financial, man-hour and materials that are not discussed in this training manual.

An activity list is a tabulation of activities to be included on a project schedule that includes:
i) The activity name
ii) An activity identifier or number
iii) A brief description of the activity

Activity attributes provide more information such as predecessors, successors, logical relationships, leads and lags, resource requirements, constraints, imposed dates, and assumptions related to the activity

Example

A project network analysis is illustrated with reference to the key information if the Table 1 below. Suppose that Company A is going to carry out a minor redesign of a product and its associated packaging. The Company intends to test a market for the redesigned product and then revise it in the light of the test market results, finally presenting the results to the Board of the company for mass production approval.

However, there are other factors to be considered such as costs and manpower but key question is: ***How long will it take to complete this project?***

After much experts discussions, they have identified the following list of separate activities together with their associated schedules (completion times) that are assumed known with certainty.

Table 1: Product Manufacture Activity List.

Activity Number	Activity Name	Completion time (Weeks)
1	Redesign product	6
2	Redesign packaging	2
3	Order and receive components for redesigned product	3
4	Order and receive material for redesigned packaging	2
5	Assemble products	4
6	Make up packaging	1
7	Package redesigned product	1
8	Test market redesigned product	6
9	Revise redesigned product	3
10	Revise redesigned packaging	1
11	Present results to the Board	1

When such a list of activities is developed should go together with making judgements as to the level of detail (timescale) to adopt. It is also important that, for clarity, keep this list to a minimum by specifying only *immediate* relationships, that is relationships involving activities that "occur near to each other in time".

Another list a part from the list of activities is the list of precedence relationships indicating activities that, because of the logic of the situation, must be finished before other activities can start.

From Table 1, one could say that:

i) Activity number 1 must be finished before activity number 3 can start.
ii) Activity 1 must be finished before activity 9 can start, but these two activities can hardly be said to have an immediate relationship (since many other activities after activity 1 need to be finished before we can start activity 9).
iii) Activities 8 and 9 would be examples of activities that have an immediate relationship (activity 8 must be finished before activity 9 can start).

c) Critical Path Method (CPM)

Critical path method (CPM)—also called critical path analysis—is a network diagramming technique used to predict total project duration. This important tool helps you combat project schedule overruns. A critical path for a project is the series of activities that determine the earliest time by which the project can be completed. It is the longest path through the network diagram and has the least amount of slack or float. Slack or float is the amount of time an activity may be delayed without delaying a succeeding activity or the project finish date.

d) Program Evaluation and Review Technique (PERT)

Program Evaluation and Review Technique (PERT)—a network analysis technique used to estimate project duration when there is a high degree of uncertainty about the individual activity duration estimates. PERT applies the critical path method (CPM) to a weighted average duration estimate.

PERT uses probabilistic time estimates—duration estimates based on using optimistic, most likely, and pessimistic estimates of activity durations—instead of one specific or discrete duration estimate, as CPM does. To use PERT, you calculate a weighted average for the duration estimate of each project activity using the following formula:

$$\text{PERT weighted average} = \frac{\text{optimistic time} + 4 * \text{most likely time} + \text{pessimistic time}}{6}$$

By using the PERT weighted average for each activity duration estimate, the total project duration estimate takes into account the risk or uncertainty in the individual activity estimates.

Example

Suppose that Mzumbe University project team in the opening case used PERT to determine the schedule for the online registration system project. Without using PERT, the duration estimate for that activity would be 10 workdays to completely design an input screen for the system. Suppose an optimistic estimate is that the input screen can be designed in eight workdays, and a pessimistic time estimate is 24 workdays.

Applying the PERT formula, you get the following:

$$\text{PERT weighted average} = \frac{8 \text{ workdays} + 4 * 10 \text{ workdays} + 24 \text{ workdays}}{6}$$
$$= 12 \text{ workdays}$$

Instead of using the most likely duration estimate of 10 workdays, the project team would use 12 workdays when doing critical path analysis. These additional two days could help the project team get the work completed on time.

The main advantage of PERT is that it attempts to address the risk associated with duration estimates. Because many projects exceed schedule estimates, PERT may help in developing schedules that are more realistic. PERT's main disadvantages are that it involves more work than CPM because it requires several duration estimates, and there are better probabilistic methods for assessing schedule risk.

e) Work Breakdown Structure

The development of a project plan is predicated on having a clear, and detailed understanding of both the tasks involved, the estimated length of time each task will take, the dependencies between those tasks, and the sequence in which those tasks have to be performed. One of the methods used to develop the list of tasks is to create what is known as a Work Breakdown Structure (WBS).

A Work Breakdown Structure (WBS) is a hierarchic decomposition or breakdown of a project or major activity into successively levels, where each level is a finer breakdown of the preceding one. In final form a WBS is very similar in structure and layout to a document outline. Each item at a specific level of a WBS is numbered consecutively (e.g. 10, 20, 30, 40, 50). Each item at the next level is numbered within the number of its parent item (e.g. 10.1, 10.2, 10,3, 10.4).

Example

Table 2. One example of the WBS

Number	Task Description
1.0	Requirements Gathering
1.1	Product design
1.1. 0	Database design
1.1.1	User interface design
1.2	Product testing
1.2.0	Integration Testing
1.2.1	User acceptance Testing
1.3	Product deployment
1.4	Product mass production
1.5	Product Marketing

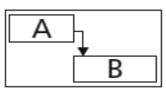
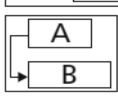
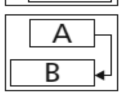
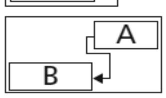
f) Activity Dependencies

Dependencies are the relationships among tasks that determine the order in which activities need to be performed. There are four (4) types of dependency relationships.

There are four types of dependencies:

i) End to start/ **Finish-to-start (FS):** (the most common dependency type)-The successor activity's start depends on the completions of the successor activity. (i.e. Activity B can't start until activity A is finished)
ii) **Start-to-start (SS)**: The start of the successor activity depends on the start of the predecessor activity. (i.e. Activity B can only start if activity A has already started).
iii) End to end/ **Finish-to-finish (FF):** The completion of the successor activity depends on the completion of the predecessor activity. (i.e. Activity B can only finish if activity A has finished).
iv) Start to end / **Start-to-finish (SF):** The completion of the successor activity depends on the start of the predecessor activity.(i.e. Activity B can't finish until activity A starts).

Figure 1: Tasks Dependencies Types

Task dependency	Example	Description
Finish-to-start (FS)		Task (B) cannot start until task (A) finishes.
Start-to-start (SS)		Task (B) cannot start until task (A) starts.
Finish-to-finish (FF)		Task (B) cannot finish until task (A) finishes.
Start-to-finish (SF)		Task (B) cannot finish until task (A) starts.

To develop a precedence relationship based on Table 2 above, aided by the fact that activities are listed in a logical/chronological order one can come up with the following list of immediate precedence relationships.

Table 3: Product Activities Precedence of Relationships

Activity Number	Activity Number
1 must be finished before	3 can start
2	4
3	5
4	6
5, 6	7
7	8
8	9
8	10
9, 10	11

Tip

The key to constructing this table is, for each activity in turn, to ask the question:
"What activities must be finished before this activity can start?"

i) activities 1 and 2 do not appear in the right hand column of the above table, this is because there are no activities which must finish before they can start, i.e. both activities 1 and 2 can start immediately
ii) two activities (5 and 6) must be finished before activity 7 can start
iii) it is plain from this table that non-immediate precedence relationships (e.g. "activity 1 must be finished before activity 9 can start") need not be included in the list since they can be deduced from the relationships already in the list.

Before starting any of the above activity, the questions asked would be:

i) What activities must be finished before this activity can start?
ii) could we complete this project in 30 weeks?
iii) could we complete this project in 2 weeks?

One answer could be, if we first do activity 1, then activity 2, then activity 3, ..., then activity 10, then activity 11 and the project would then take the sum of the activity completion times, 30 weeks.

"What is the minimum possible time in which we can complete this project?

Later, we shall see how the network analysis diagram/picture when well-constructed helps to answer questions like this.

Network analysis is likely to be of most value where projects are:
i) Complex. That is, they contain many related and interdependent activities.
ii) Large, that is many types of facilities (high capital investment, many personnel, etc) are involved.
iii) Where restrictions exists. That is where projects have to be completed within stipulated time or cost limits, or where some or all of the resources are limited.

Such projects might be for example:
i) The construction of a transport link;
ii) Preparing a dinner party (organizing a wedding)
iii) Launching a new product. Etc.

i) Activity Network Diagram

a) Definition

Activity Network Diagram depicts the flow of various activities involved in a project in the order of time. It depicts clearly the dependencies existing between activities (i.e.: the activities that need to be completed to proceed to any individual activity) and activities that can be performed together without affecting each other. In other terms: Precedence and Parallel processes/activities. Activity Network Diagram is a graphical way to view tasks, dependencies, and the critical path of your project. Boxes (or nodes) represent tasks, and dependencies show up as lines that connect those boxes.

Activity Network Diagrams started out as an engineering and construction project management tool. An Activity Network Diagram helps to find out the most efficient sequence of events needed to complete any project. It enables you to create a realistic project schedule by graphically showing

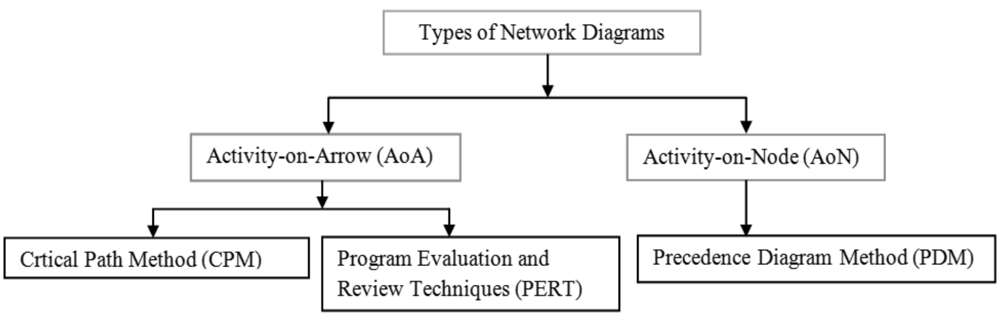
1. the total amount of time needed to complete the project
2. the sequence in which tasks must be carried out
3. which tasks can be carried out at the same time
4. which are the critical tasks that you need to keep an eye on.

There are two types of network diagrams.

i) Activity on Arrow diagram or AOA or Arrow Diagram
ii) Activity on Node diagram or Precedence diagram method(PDM)

An Activity Network Diagram is also called an Arrow Diagram (because the pictorial display has arrows in it) or a PERT (Program Evaluation Review Technique) Diagram and it is used for identifying time sequences of events that are pivotal to objectives. Program Evaluation and Review Technique (**PERT**) and Critical Path Method (**CPM**) help managers to plan the timing of projects involving sequential activities. **PERT**/**CPM** charts identify the time required to complete the activities in a project, and the order of the steps.

Figure 2: Network Diagram Composition



As mentioned in the Figure2, CPM/CPA (Critical Path Method/Critical Path Analysis) and PERT (Program Evaluation and Review Technique) fall under Arrow Network Diagram. PDM falls under Activity Network diagram or AON.

Tip

i) The Program Evaluation and Review Technique (PERT) is a project management technique or tool which is suitable for projects that have unpredictable activities while the Critical Path Method (CPM) is a project management tool which is suitable for projects that have predictable activities.

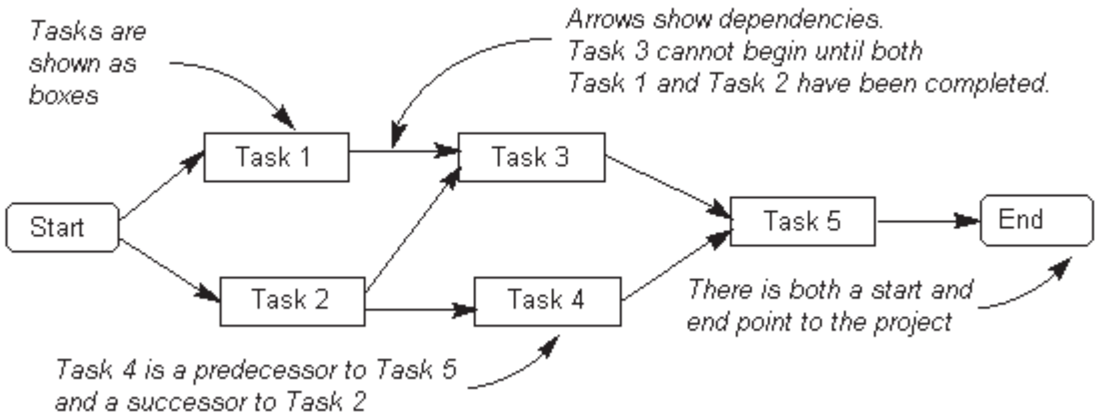
ii) CPM uses a single estimate for the time that a project can be completed while PERT uses three estimates for the time that it can be completed.

iii) CPM is a deterministic project management tool while PERT is a probabilistic project management tool.

iv) CPM allows project management planners to determine which aspect of the project to sacrifice when a trade-off is needed in order to complete the project while PERT does not.

Example of Activity Network Diagram

Figure 3: Project Activity Network Diagram (Activity-on-Node)



The most key features of the Activity Network diagram are:

i) Boxes and arrows- The Activity Network diagram displays interdependencies between tasks using boxes and arrows.
ii) Arrows pointing into a task box come from its predecessor tasks, which must be completed before the task can start.
iii) Arrows pointing out of a task box go to its successor tasks, which cannot start until at least this task is complete.

How to Create an Activity Network Diagram:

1. First list down all the activities involved in the project
2. Find out the chronological order of the activities (i.e.: Preceding activities that should necessarily be completed before a given activity)
3. Find out tasks that can be executed simultaneously.

Imagine a project with the below activities:

Activity
T1
T2
T3
T4
T5
T6

Now let us assume that the preceding activities for each activity are as below:

Activity	Preceding Task
T1	None
T2	T1
T3	T1
T4	T2
T5	T3
T6	T4

4. To draw an Activity Network Diagram, one must know that the Activity Network Diagram is made of Nodes and Arrows. Arrows depict an activity, whereas the nodes depict the start and end of an activity. For example; Activity T1 will have two nodes A and B which depict the start and end of the activity T1. The notation A-B also indicates the task T1.
5. The next activities T2, T3 require T1 to be completed. Therefore two arrows emerge from the node B simultaneously. They can be depicted as B-C, B-D.
6. Similarly, T4 is dependent on T2 and T6 on T3 respectively. Therefore a single node is drawn from C and D, which are C-E and D-F depicting T4 and T6 respectively.
7. And T5 is dependent on T4 and hence E-F is drawn. (Since there are no further activities, the nodes converge to a common end point.

Activity	Preceding Task	Node	Preceding Node
T1	**None**	**A-B**	**None**
T2	**T1**	**B-C**	**A-B**
T3	**T1**	**B-D**	**A-B**
T4	**T2**	**C-E**	**B-C**
T5	**T3**	**E-F**	**C-E**
T6	**T4**	**D-F**	**B-D**

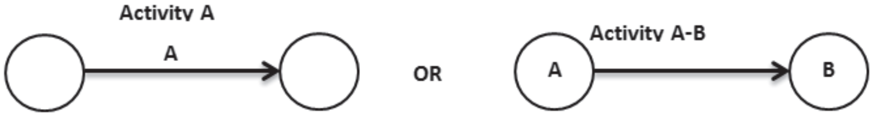
Note: There should be only a single Start and End Nodes for a project. Each task should be given unique node

a) Network Diagram Using CPM / PERT

The network diagram using CPM/PERT use **Activity-On-Arrow (AOA)** network diagram in which the arrows are used to represent activities. The nodes represent activity dependencies. Any activity coming into a node is a predecessor to any activity leaving the node.

This method uses only finish to start relationship between activities. This may use dummy activities that should be inserted just to show dependencies between activities. The dummy activity do not need work or take any time.

Figure 4: Activity-On-Arrow (AOA) Presentation



As shown in the Figure 4, there are two ways we can denote activity on arrow(AOA) activity. Either the name will be on arrow as shown in Figure 4 left side or activity will be named as shown in the Figure 4 right side. This method requires two letter identifiers to name a single activity (e.g. A-B).

The characteristics of the Activity on Arrow (A-O-A) or (A-O-L) are:-

i) It is composed of arrows and nodes. The arrows represent the activities and nodes represent the events.
ii) Each activity carries a brief description usually printed on the logical diagram, the activity name or symbol and the time duration.
iii) At present, this method seems to be the most popular method and it was the first method to be introduced, developed and computerized

Essentially, six steps are common to both AoA and AoN techniques. The procedures are:
i) Define the Project and all of it's significant activities or tasks. The Project (made up of several tasks) should have only a single start activity and a single finish activity.
ii) Develop the relationships among the activities. Decide which activities must precede and which must follow others.
iii) Draw the "Network" connecting all the activities. Each Activity should have unique event numbers. Dummy arrows are used where required to avoid giving the same numbering to two activities.
iv) Assign time and/or cost estimates to each activity
v) Compute the longest time path through the network. This is called the critical path.
vi) Use the Network to help plan, schedule, monitor and control the project.

Steps in CPM Project Planning

i) Specify the individual activities: From the Work Breakdown Structure, a listing can be made of all the activities in the project. This listing can be used as the basis for adding sequence and duration information in later steps.
ii) Determine the sequence of those activities: Some activities are dependent upon the completion of others. A listing of the immediate predecessors of each activity is useful for constructing the CPM Network diagram.
iii) Draw a network diagram : Once the activities and their sequencing have been defined, the CPM diagram can be drawn (usually as an activity on node (AON) network)
iv) Estimate the completion time for each activity: The time required to complete each activity can be estimated using experience or the estimates of knowledgeable persons. CPM is a deterministic model that does not take into account variation in the completion time, so only one number can be used for an activity's time estimate.
v) Identify the critical path (the longest path through the network): The critical path is the longest-duration path through the network.
vi) Update the CPM diagram as the project progresses: As the project progresses, the actual task completion times will be known and the diagram can be updated to include this information. A new critical path may emerge, and structural changes may be made in the network if project requirements change.

Consider the list of four activities for making a simple product:

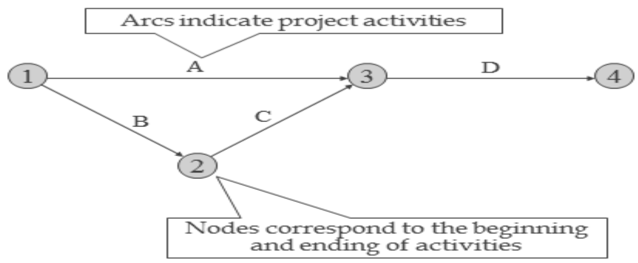
Activity	Description	Immediate Predecessors
A	Buy Plastic Body	-
B	Design Component	-
C	Make Component	B
D	Assemble product	A, C

Table 3 above indicates that:

- Immediate predecessors for a particular activity are the activities that, when completed, enable the start of the activity in question
- Can start work on activities A and B anytime, since neither of these activities depends upon the completion of prior activities.
- Activity C cannot be started until activity B has been completed
- Activity D cannot be started until both activities A and C have been completed.
- The graphical representation (Figure 5 is referred to as the PERT/CPM network

Example 1-

Figure 5: A Simple Network of Four Activities



Example 2

Develop the network for a project with following activities and immediate predecessors.

Activity	Immediate predecessors
A	-
B	-
C	B
D	A, C
E	C
F	C
G	D, E, F

Try to do for the first five (A, B, C,D,E) activities

Figure 6: Network of first five activities

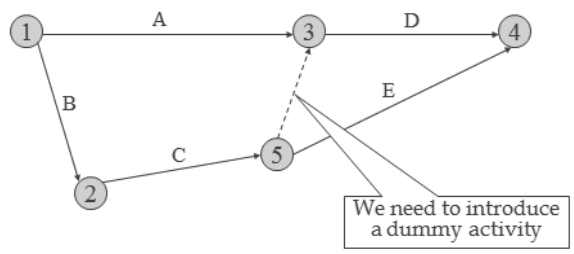
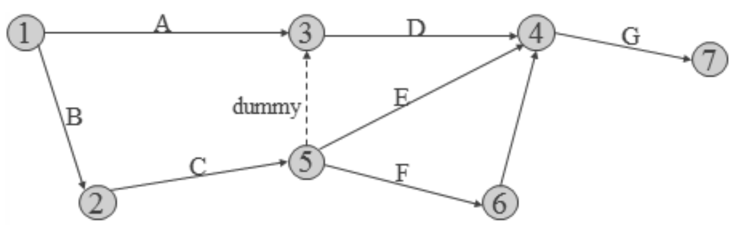


Figure 7: Network of Seven Activities



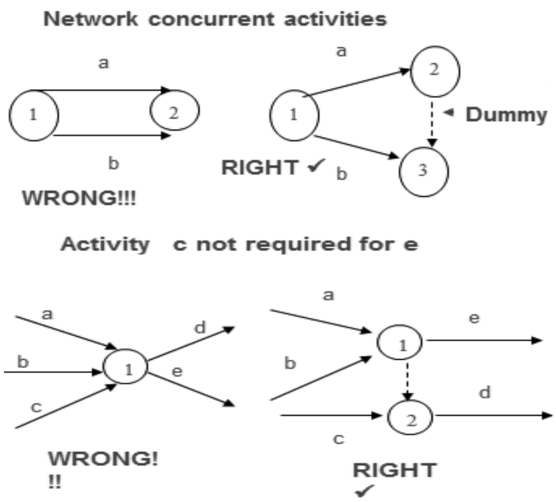
Reading Figure 7, will give the following information:

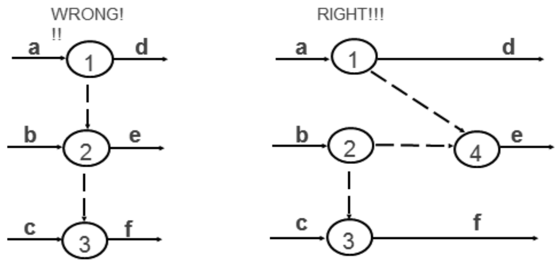
i) Note how the network correctly identifies D, E, and F as the immediate predecessors for activity G.

ii) Dummy activities is used to identify precedence relationships correctly and to eliminate possible confusion of two or more activities having the same starting and ending nodes

iii) Dummy activities have no resources (time, labor, machinery, etc) – purpose is to PRESERVE LOGIC of the network

Examples of The Use of Dummy activity





- a precedes d.
- a and b precede e,
- b and c precede f (a does not precede f)

The Key Concept used by CPM/PERT is that a small set of activities, which make up the longest path through the activity network control the entire project. If these "critical" activities could be identified and assigned to responsible persons, management resources could be optimally used by concentrating on the few activities which determine the fate of the entire project. Non-critical activities can be re-planned, rescheduled and resources for them can be reallocated flexibly, without affecting the whole project.

Five useful questions to ask when preparing an activity network are:

i) Is this a Start Activity?
ii) Is this a Finish Activity?
iii) What Activity Precedes this?
iv) What Activity Follows this?
v) What Activity is Concurrent with this?

Some activities are serially linked. The second activity can begin only after the first activity is completed. In certain cases, the activities are concurrent, because they are independent of each other and can start simultaneously. This is especially the case in organisations that have supervisory resources so that work can be delegated to various departments that will be responsible for the activities and their completion as planned.
Consider a Project for manufacturing a product with only Four major activities.

Key Differences Between PERT and CPM

PERT	CPM
PERT is a project management technique, whereby planning, scheduling, organizing, coordinating and controlling of uncertain activities is done.	CPM is a statistical technique of project management in which planning, scheduling, organizing, coordination and control of well-defined activities takes place.
PERT is a technique of planning and control of time	Unlike CPM, which is a method to control costs and time.
While PERT is evolved as research and development project.	CPM evolved as construction project.
PERT is set according to events	Conversely, PERT uses probabilistic model.
There are three times estimates in PERT i.e. optimistic time (to), most likely time ™,	there is only one estimate in CPM.

pessimistic time (tp)	
PERT technique is best suited for a high precision time estimate	CPM is appropriate for a reasonable time estimate.
PERT deals with unpredictable activities.	CPM deals with predictable activities.
PERT is used where the nature of the job is non-repetitive.	CPM involves the job of repetitive nature.
No Demarcation between critical and non-critical activities.	There is a demarcation between critical and non-critical activities in CPM.
PERT is best for research and development projects	CPM is for non-research projects like construction projects.
Crashing is a compression technique applied to CPM, to shorten the project duration, along with least additional cost.	The crashing concept is not applicable to PERT.

Benefits of CPM / PERT Network

i) Consistent framework for planning, scheduling, monitoring, and controlling project.
ii) Shows interdependence of all tasks, work packages, and work units.
iii) Helps proper communications between departments and functions.
iv) Determines expected project completion date.
v) Identifies so-called critical activities, which can delay the project completion time.
vi) Identified activities with slacks that can be delayed for specified periods without penalty, or from which resources may be temporarily borrowed
vii) Determines the dates on which tasks may be started or must be started if the project is to stay in schedule.
viii) Shows which tasks must be coordinated to avoid resource or timing conflicts.
ix) Shows which tasks may run in parallel to meet project completion date

NETWORK ANALYSIS AND PROJECT SCHEDULING

10

Study Guide 10 -B: PROJECT COST, CRITICAL PATHS DURATIONS , CRITICAL PATH ACTIVITIES AND CRASHING OF ACTIVITY

Get Through Intro

The Critical Path Method (CPM) was developed in the 1950s by the US Navy. Originally, the critical path method considered only logical dependencies between terminal elements. Since then, it has been expanded to allow for the inclusion of resources related to each activity, through processes called activity-based resource assignments and resource levelling.

The Critical Path Method or Critical Path Analysis is a mathematically based algorithm for scheduling a set of project activities. It is an important tool for effective project management that is commonly used with all forms of projects, including construction, software development, research projects, product development, engineering, and plant maintenance, among others. Any project with interdependent activities can apply this method of scheduling. The essential technique for using CPM is to construct a model of the project that includes the following:

i) A list of all activities required to complete the project (also known as Work Breakdown Structure)

ii) The time (duration) that each activity will take to completion

iii) The dependencies between the activities.

CPM provides the following benefits:

i) Provides a graphical view of the project

ii) Predicts the time required to complete the project.

iii) Shows which activities are critical to maintaining the schedule and which are not.

CPM models the events and activities of a project as a network. Activities are depicted as nodes on the network and events that signify the beginning or ending of activities and are depicted as arcs or lines between the nodes.

Learning Outcomes

a) Determine the duration times of uncertain activities.
b) Determine the duration of a project.
c) Identify the critical path activity.
d) Determine the cost of the project.
e) Crash activities.

The Critical Path Method (CPM) calculates

i) The longest path of planned activities to the end of the project
ii) The earliest and latest that each activity can start and finish without making the project longer Project Scheduling with activity time

Determines "critical" activities (on the longest path) by prioritizing activities for the effective management and to shorten the planned critical path of a project by:

i) Pruning critical path activities
ii) "Fast tracking" (performing more activities in parallel)
iii) "Crashing the critical path" (shortening the durations of critical path activities by adding resources).

The critical path can be identified by determining the following four parameters for each activity:

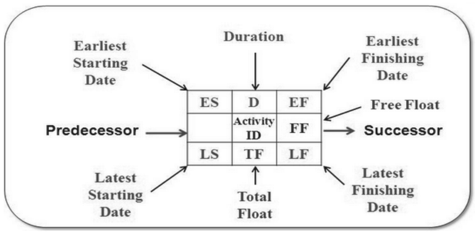
i) ES – earliest start time: the earliest time at which an activity can begin given that its predecessor activities must be completed first.
ii) EF – earliest finish time, equal to the earliest start time for the activity plus the time required to complete the activity.
iii) LF – latest finish time: the latest time at which an activity can be completed without delaying the project.
iv) LS – latest start time, equal to the latest finish time minus the time required to complete the activity.

The slack or float time for an activity is the time between the earliest and latest start time, or between the earliest and latest finish time. Slack is the amount of time that an activity can be delayed past its earliest start or earliest finish without delaying the project.

The critical path is the path through the project network in which none of the activities have slack, that is, the path for which LS=ES and LF=EF for all activities in the path. A delay in the critical path delays the project. Similarly, to accelerate the project it is necessary to reduce the total time required for the activities in the critical path.

The CPM Approach:

i) Break project into operations necessary for completion. Define the required tasks and put them down in an ordered (sequenced) list.
ii) Determine sequential relationship of operations. Create a flowchart or other diagram showing each task in relation to the others.
iii) Create time estimates for each operation.
iv) Determine earliest possible start date, earliest possible finish date, latest start & finish.
v) Determine "free float" and "total float" Identify the critical and non-critical relationships (paths) among tasks.
vi) Establish time-cost relationship
vii) Establish scheduling variations. Determine the expected completion or execution time for each task. Determine most favorable balance between time-cost.
viii) Normal Start – normal time, least cost.
ix) All-Crash Start – least time, higher cost
x) Locate or devise alternatives (backups) for the most critical paths.



Project Scheduling with activity time

Table 1: Activities for Project M

Activity	Immediate predecessors	Completion Time (week)
A	-	5
B	-	6
C	A	4
D	A	3
E	A	1
F	E	4
G	D, F	14
H	B,C	12
I	G, H	2
Total		51

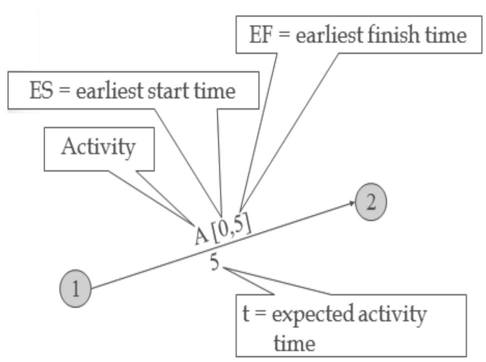
This information indicates that the total time required to complete activities is 51 weeks. However, we can see from the network that several of the activities can be conducted simultaneously (A and B, for example).

Earliest start & earliest finish time (ES and EF)
We are interested in the longest path through the network, i.e., the critical path. Starting at the network's origin (node 1) and using a starting time of 0, we compute an ***Earliest Start (ES)*** and ***Earliest Finish (EF)*** time for each activity in the network.

The expression EF = ES + t can be used to find the earliest finish time for a given activity.
For example, for activity A, ES = 0 and t = 5; thus the earliest finish time for activity A is

EF = 0 + 5 = 5

Figure 1: Arc with ES & EF time



Activity, duration, ES, EF, LS, LF

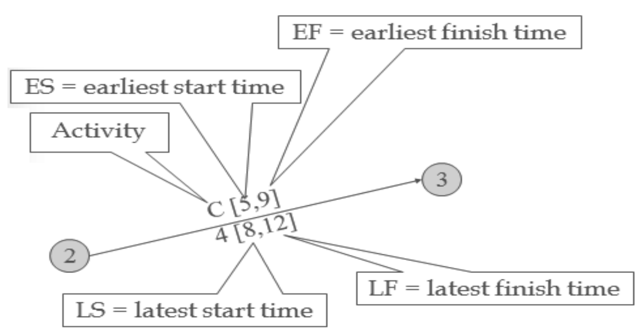
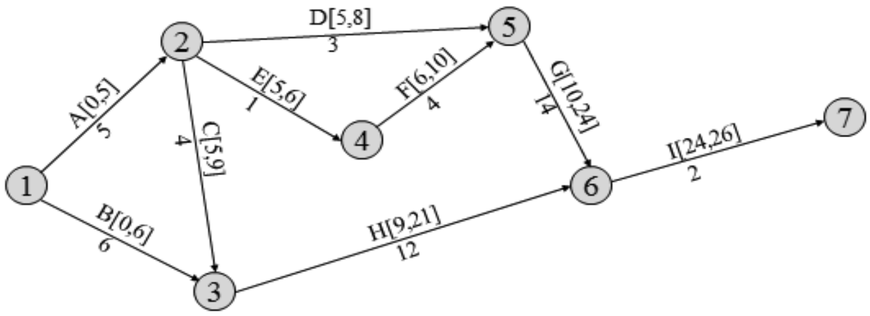


Figure 2: Network with ES & EF time

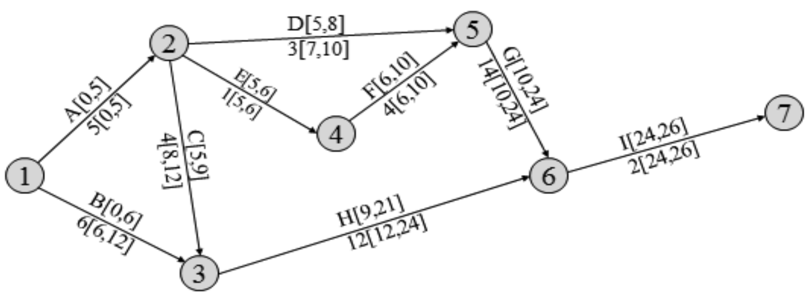


Earliest start time rule:
The earliest start time for an activity leaving a particular node is equal to the largest of the earliest finish times for all activities entering the node.

Latest start & latest finish time
To find the critical path we need a backward pass calculation.

i) Starting at the completion point (node 7) and using a latest finish time (LF) of 26 for activity I, we trace back through the network computing a latest start (LS) and latest finish time for each activity
ii) The expression LS = LF – t can be used to calculate latest start time for each activity. For example, for activity I, LF = 26 and t = 2, thus the latest start time for activity I is **LS = 26 – 2 = 24**

Figure 3: Network with LS & LF time



Latest finish time rule:
The latest finish time for an activity entering a particular node is equal to the smallest of the latest start times for all activities leaving the node.

Slack or Free Time or Float

Slack is the length of time an activity can be delayed without affecting the completion date for the entire project.
For example, slack for C = 3 weeks, i.e Activity C can be delayed up to 3 weeks (start anywhere between weeks 5 and 8).

Figure 4: Slack for Activity C

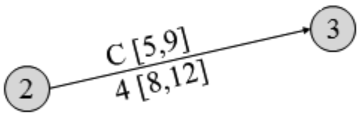


Figure 5: Slack or Free Time or Float

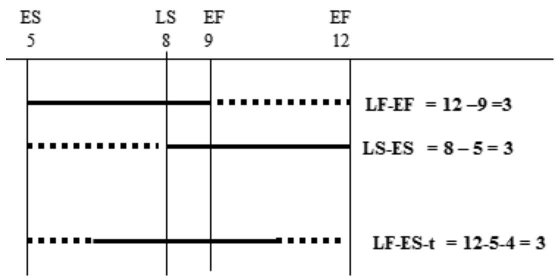


Table 3: Activity schedule for Project M (**Table xxb)**

Activity	**Earliest start (ES)**	**Latest start (LS)**	**Earliest finish (EF)**	**Latest finish (LF)**	**Slack (LS-ES)**	**Critical path**
A	0	0	5	5	0	YES
B	0	6	6	12	6	
C	5	8	9	12	3	
D	5	7	8	10	2	
E	5	5	6	6	0	YES
F	6	6	10	10	0	YES
G	10	10	24	24	0	YES
H	9	12	21	24	3	
I	24	24	26	26	0	YES

Important Questions

i) **What is the total time to complete the project?**
   – 26 weeks if the individual activities are completed on schedule.

ii) **What are the scheduled start and completion times for each activity?**
   – ES, EF, LS, LF are given for each activity.

iii) **What activities are *critical* and must be completed as scheduled in order to keep the project on time?**
   – Critical path activities: A, E, F, G, and I.

iv) **How long can *non-critical* activities be delayed before they cause a delay in the project's completion time**
   – Slack time available for all activities are given.

Importance of Float (Slack) and Critical Path

i) Slack or Float shows how much allowance each activity has, i.e how long it can be delayed without affecting completion date of project

ii) Critical path is a sequence of activities from start to finish with zero slack. Critical activities are activities on the critical path.

iii) Critical path identifies the minimum time to complete project

iv) If any activity on the critical path is shortened or extended, project time will be shortened or extended accordingly

v) So, a lot of effort should be put in trying to control activities along this path, so that project can meet due date. If any activity is lengthened, be aware that project will not meet deadline and some action needs to be taken.

vi) If can spend resources to speed up some activity, do so only for critical activities.

vii) Don't waste resources on non-critical activity, it will not shorten the project time.

viii) If resources can be saved by lengthening some activities, do so for non-critical activities, up to limit of float.

ix) Total Float belongs to the path

PERT For Dealing With Uncertainty

Times can be estimated with relative certainty, confidence. For many situations, this is not possible, e.g Research, development, new products and projects etc.

Use 3 time estimates

m= most likely time estimate, mode.

a = optimistic time estimate,

b = pessimistic time estimate, and

Expected Value (TE) = (a + 4m + b) /6

Variance (V) = ((b – a) / 6) 2

Std Deviation (δ) = SQRT (V)

Table 3: **Precedencies And Project Activity Times**

Activity	**Immediate predecessors**	**Optimistic Time**	**Most likely time**	**Pessimistic time**	**EX TE**	**Var V**	**S.Dev δ**
A	-	10	22	22	20	4	2
B	-	20	20	20	20	0	0
C	-	4	10	16	10	4	2
D	A	2	14	32	15	25	5
E	B, C	8	8	20	10	4	2
F	B, C	8	14	20	14	4	2
G	B, C	4	4	4	4	0	0
H	c	2	12	16	11	5.4	2.32
I	G,H	6	16	38	18	28.4	5.33
J	D, E	2	8	14	8	4	2

Figure 6: The complete network (a)

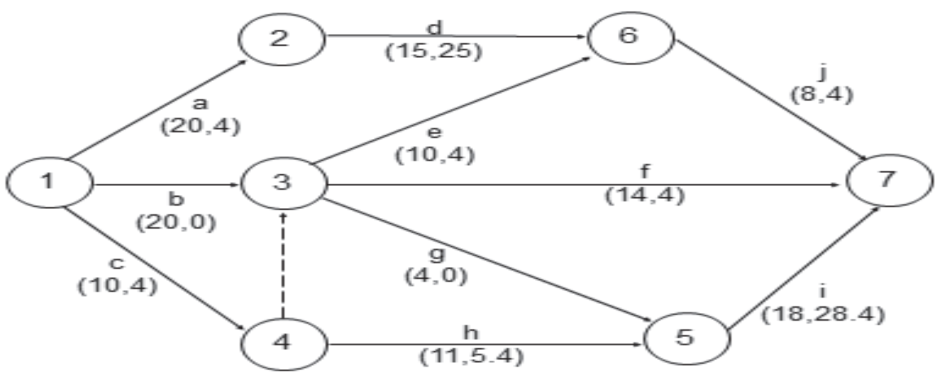


Figure 7: The complete network (b)

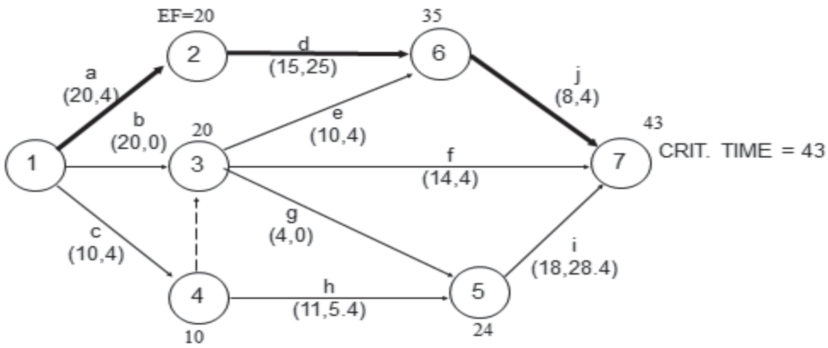


Table 4: Critical Path Analysis (PERT)

Activity	LS	ES	Slacks	Critical?
A	**0**	**0**	**0**	**YES**
B	1	0	1	
C	4	0	4	
D	**20**	**20**	**0**	**YES**
E	25	20	5	
F	29	20	9	
G	21	20	1	
H	14	10	4	
I	25	24	1	
J	**35**	**35**	**0**	**YES**

Assume, Project management promised to complete the project in the fifty days. What are the chances of meeting that deadline?

Calculate Z, where **$Z = (D-S) / \sqrt{V}$**

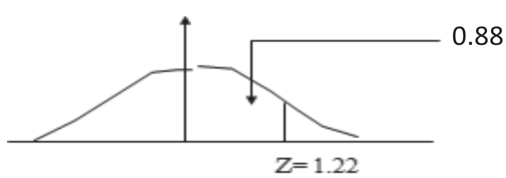
Example

D = 50; S(Scheduled date) = 20+15+8 =43; V = (4+25+4) =33

Z = (50 – 43) / 5.745

= 1.22 standard deviations.

The probability value of Z = 1.22, is 0.888



What deadline are you 95% sure of meeting.

Z value associated with 0.95 is 1.645

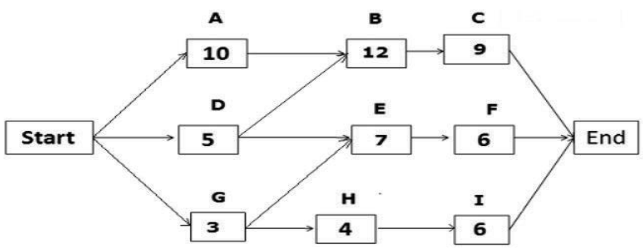
D = S + 5.745 (1.645)

= 43 + 9.45

= 52.45 days

Thus, there is a 95 percent chance of finishing the project by 52.45 days

Example:
Based on the below network diagram, identify the total paths, critical path, and float for each path.



The above network diagram has five paths; the paths and their duration are as follows:

i) Start -> A -> B -> C-> End, duration: 31 days.
ii) Start ->D -> E ->F -> End, duration: 18 days.
iii) Start -> D -> B -> C -> End, duration: 26 days.
iv) Start -> G ->H ->I -> End, duration: 13 days.
v) Start -> G -> E ->F -> End, duration: 16 days.

Since the duration of the first path is the longest, it is the critical path. The float on the critical path is zero. The float for the second path "Start ->D -> E ->F -> End" = duration of the critical path – duration of the path "Start ->D -> E ->F -> End"

= 31 – 18 = 13

Hence, the float for the second path is 13 days.

Using the same process, we can calculate the float for other paths as well.

Float for the third path = 31 – 26 = 5 days.
Float for the fourth path = 31 – 13 = 18 days.
Float for the fifth path = 31 – 16 = 15 days.

Calculate Early Start (ES), Early Finish (EF), Late Start (LS), and Late Finish (LF)
We have identified the critical path, and the duration of the other paths, it's time to move on to more advanced calculations, Early Start, Early Finish, Late Start and Late Finish.

- **Calculating Early Start (ES) and Early Finish (EF)**

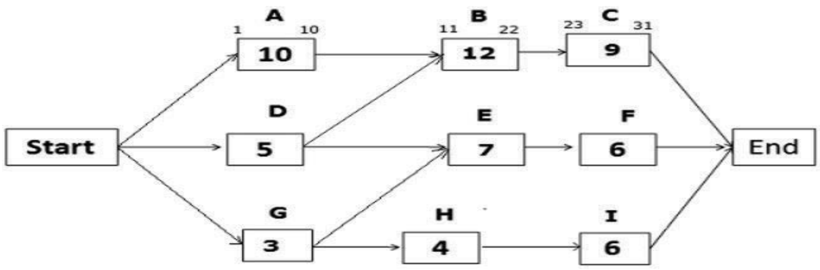
To calculate the Early Start and Early Finish dates, we use forward pass; we will start from the beginning and proceed to the end.

Early Start (ES) for the first activity on any path will be 1, because no activity can be started before the first day. The start point for any activity or step along the path is the end point of the predecessor activity on the path plus one.

The formula used for calculating Early Start and Early Finish dates.

i) Early Start of the activity = Early Finish of predecessor activity + 1
ii) Early Finish of the activity = Activity duration + Early Start of activity – 1

Early Start and Early Finish Dates for the path Start -> A -> B -> C -> End



Early Start of activity A = 1 *(Since this is the first activity of the path)*

Early Finish of activity A = ES of activity A + activity duration – 1 = 1 + 10 – 1 = 10

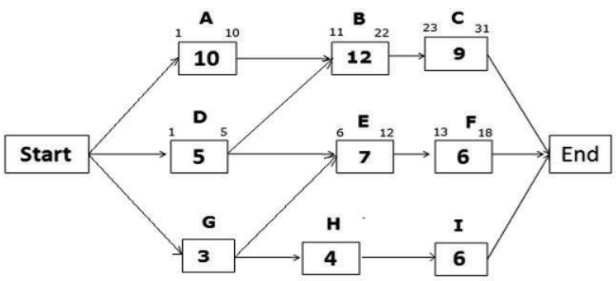
Early Start of activity B = EF of predecessor activity + 1 = 10 +1 = 11

Early Finish of activity B = ES of activity B + activity duration – 1 = 11 + 12 – 1 = 22

Early Start of activity C = EF of predecessor activity + 1 = 22 +1 = 23

Early Finish of activity C = ES of activity C + activity duration – 1 = 23 + 9 – 1 = 31

Early Start and Early Finish Dates for the path Start -> D -> E -> F -> End



Early Start of activity D = 1 *(Since this is the first activity of the path)*

Early Finish of activity D = 1 + 5 – 1 = 5

Early Start of activity E = EF of predecessor activity + 1

Since the Activity E has two predecessor activities, which one will you select? You will select the activity with the greater Early Finish date. Early Finish of activity D is 5, and Early Finish of activity G is 3 *(we will calculate it later)*.

Therefore, we will select the Early Finish of activity D to find the Early Start of activity E.

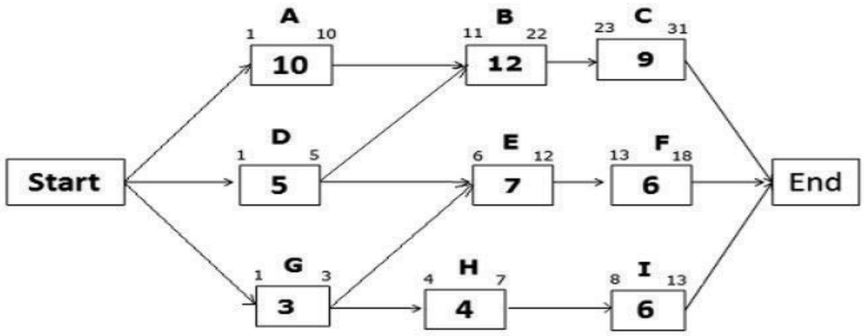
Early Start of activity E = EF of predecessor activity + 1 = 5 + 1 = 6

Early Finish of activity E = 6 + 7 – 1 = 12

Early Start of activity F = 12 + 1 = 13

Early Finish of activity F = 13 + 6 -1 = 18

Early Start and Early Finish Dates for the path Start -> G -> H -> I -> End



Early Start of activity G = 1 *(Since this is the first activity of the path)*
Early Finish of activity G = 1 + 3 – 1 = 3

Early Start of activity H = 3 + 1 = 4
Early Finish of activity H = 4 + 4 – 1 = 7

Early Start of activity I = 7 +1 = 8
Early Finish of activity I = 8 + 6 – 1 = 13

Calculating Late Start (LS) and Late Finish (LF)

We have calculated Early Start and Early Finish dates of all activities. Now it is time to calculate the Late Start and Late Finish dates.

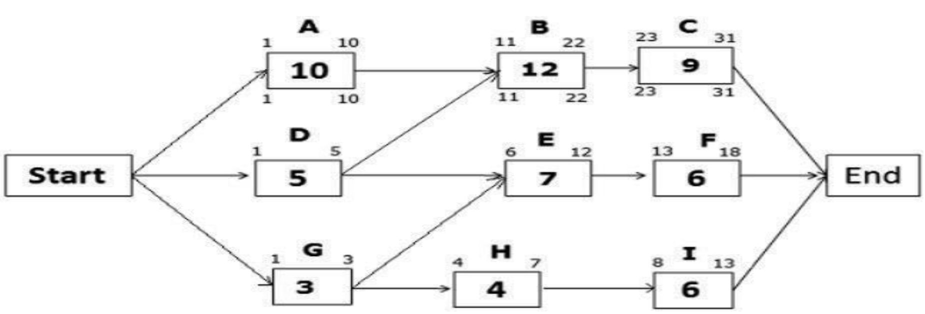
Late Finish of the last activity in any path will be the same as the Last Finish of the last activity on the critical path, because you cannot continue any activity once the project is completed.

The formula used for Late Start and Late Finish dates are:

i) Late Start of Activity = Late Finish of activity – activity duration + 1
ii) Late Finish of Activity = Late Start of successor activity – 1

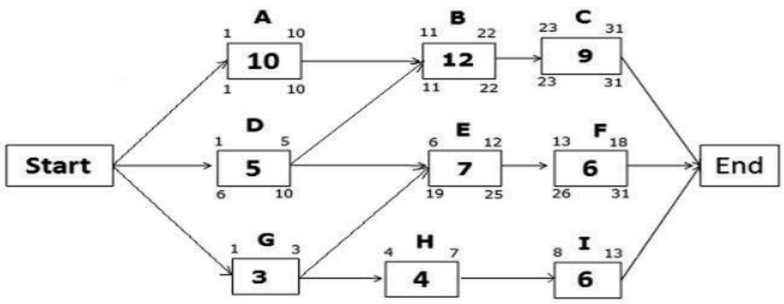
To calculate the Late Start and Late Finish, we use backward pass; i.e. we will start from the last activity and move back towards the first activity.

Late Start and Late Finish Dates for the path Start -> A -> B -> C -> End



On a critical path, Early Start, and Early Finish dates will be the same as Late Start and Late Finish dates.

Late Start and Late Finish Dates for the path Start -> D -> E -> F -> End



Late Finish of activity F = 31 *(because you cannot allow any activity to cross the project completion date)*

Late Start of activity F = LF of activity F – activity duration + 1 = 31 – 6 +1 = 26

Late Finish of activity E = LS of successor activity – 1 = LS of activity F – 1 = 26 – 1 = 25

Late Start of Activity E = LF of activity E – activity duration + 1 = 25 – 7 + 1 = 19

Late Finish of activity D = LS of successor activity – 1

If you look at the network diagram, you will notice that activity D has two successor activities, B and E. So, which activity will you select?

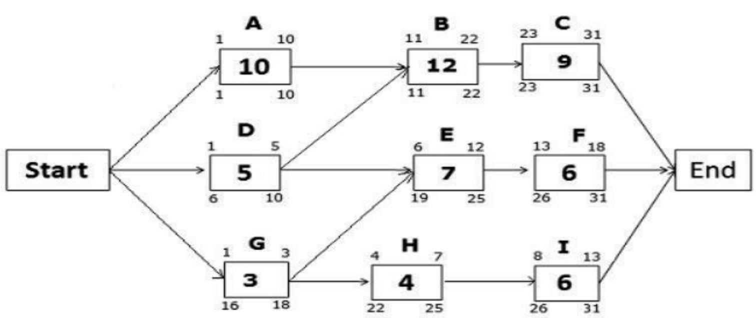
You will select the activity with the earlier(least) Late Start date. Here, Late Start of activity B is 11, and Late Start of activity E is 19.

Therefore, you will select activity B which has the earlier Late Start date. Hence,

Late Finish of activity D = LS of activity B – 1   = 11 – 1 = 10

Late Start of Activity D = LF of activity D – activity duration + 1 10 – 5 + 1 = 6

Late Start and Late Finish Dates for the path Start -> G -> H -> I -> End



Late Finish of activity I = 31 *(because you cannot allow any activity to cross the project completion date)*
Late Start of activity I = 31 – 6 + 1 = 26

Late Finish of activity H = 26 – 1 = 25
Late Start of activity H = 25 – 4 + 1 = 22

Late Finish of Activity G = 19 – 1= 18 *(we will choose the late start of activity E, not activity H, because the Late Start of activity E is earlier than the Late Start of activity H)*

Late Start of activity G = 18 – 3 + 1 = 16

b) **Total Float and Free Float**

Total float: is the amount of time an activity can be delayed without delaying the project completion date. On a critical path, the total float is zero. Total float is often known as the slack.

You can calculate the total float by subtracting the Early Start date of an activity from its Late Start date (Late Start date – Early Start date), or Early Finish date from its Late Finish date (Late Finish date – Early Finish date).

Free float: is the amount of time an activity can be delayed without delaying the Early Start of its successor activity. You can calculate the free float by subtracting the Early Finish date of the activity from the Early Start date of next activity (ES of next Activity – EF of current Activity).

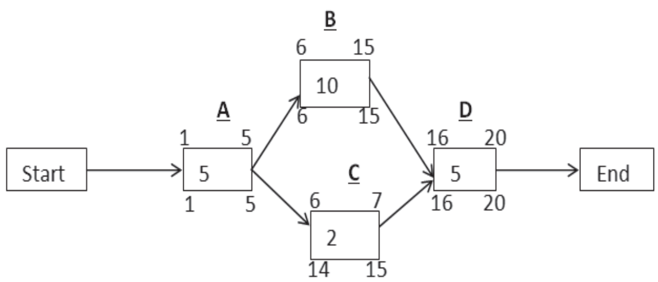
Keep in mind that if two activities are converging to a single activity, only one of these two activities may have free float.

Example

In the network diagram, we have two paths:

i) The first path is A->B->D with 20 days' duration, and
ii) The second path is A->C->D with 12 days' duration.

Obviously, the path A->B->D is the critical path because it has the longest duration.



Calculating the Total Float

As we can see, the given diagram has only two paths: path A->B->D and path A->C->D.

The path A->B->D is a critical path; therefore, it will not have a total float.

Since the path A->C->D is a non-critical path, it can have a total float.

You have two methods to calculate the total float. In the first method, you subtract the duration of the non-critical path from the critical path.

In the second method, you find the total float for any activity by subtracting the Early Start date from the Late Start date (LS – ES), or subtracting the Early Finish date from the Late Finish date (LF – EF) on any activity.

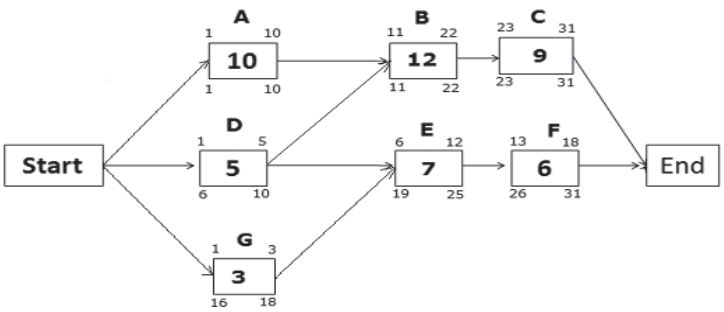
Total Float- Method ---I
Total float = duration of the critical path – duration of the non-critical path
= (duration of the path A->B->D) – (duration of the path A->C->D)
= 20 – 12
= 8
Hence, the total float is 8 days.

Total Float- Method ---I
On the path A->C->D, Activity A and D lie on the critical path; therefore, they will not have a total float. Only Activity C can have a total float.
As stated earlier, we can calculate the total float by using either finish dates or start dates. Here, I will show you both ways to find it.
First, we will go with the Late Finish and Early Finish dates:
Total float for Activity C = (LF of Activity C – EF of Activity C)
= 15 – 7 = 8
Now, the second formula:
Total float for Activity C = (LS of Activity C – ES of Activity C)
= 14 – 6 = 8
As you can see, both durations are the same, which means both formulas will provide you with the same result.

Calculating the Free Float
From the figure, you can see that only Activity C can have a free float, because other activities are lying on the critical path.
Let's find it.
Free float of Activity C = ES of next Activity – EF of Activity C – 1
= 16 – 7 – 1 = 8
Hence, the free float for activity C is 8 days.

Example

For the below given network diagram, identify which activities can have a free float and calculate the free and total float for those activities, considering duration in days.



Solution

We know that, Free float = ES of next Activity – EF of current Activity

In the above diagram, Activity G can have the free float because Activity D and G are converging on one common activity.

Activity D will not have a free float because its successor Activity E is starting on next day of completing of Activity D.

Free Float for Activity G

Free float of Activity G = Early Start of Activity E – Early Finish of Activity G – 1
= 6 – 3 – 1 = 2

Total Float for Activity G

Total float for Activity G = Late Finish of Activity G – Early Finish of Activity G
= 18 – 3 = 15

You can see here that the free float for Activity G is 2 days, and the total float is 15 days. Both are different.

c) **Precedence Diagram Method (PDM)**

Precedence Diagram Method (PDM) nodes or boxes are used to represent activities. Arrows show activity dependencies as depicted in Figure 4. The characteristics of the Activity on node (A-o-N) or Precedence diagrams are:-

i) In A-O-N networks, the nodes represent the activities and the arrows, their interdependencies or precedence relationships.
ii) Nodes are usually represented by squares or rectangles, but circles and other convenient geometrical shapes may also be used.
iii) Activity number and description are written within the boxes representing the nodes.
iv) Length and direction of the arrows have no significance as they indicate only the dependency of one activity on another.

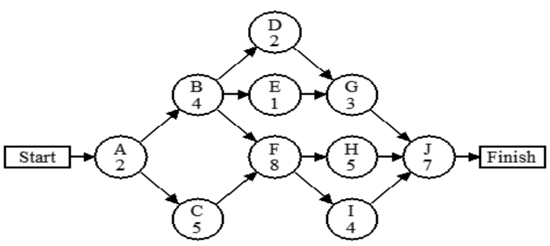
Figure 8: Activity on node (A-o-N)

There are four types of dependencies between the activities.

i) Finish to Start: An activity must finish before the successor activity can start.
ii) Start to Start: An activity must start before the successor activity can start
iii) Finish to Finish: An activity must finish before the successor activity can finish
iv) Start to Finish: An activity must start before the successor activity can finish. This type is rarely used.

Example

Consider the following project network then determine the critical path and the project duration.



The critical path is A–C–F–H–J with a completion time of 27 days.

Activity	Duration	Earliest Start	Latest Start	Earliest Finish	Latest Finish	Total Slack	On Critical Path?
A	2	0	0	2	2	0	Yes
B	4	2	3	6	7	1	No
C	5	2	2	7	7	0	Yes
D	2	6	15	8	17	9	No
E	1	6	16	7	17	10	No
F	8	7	7	15	15	0	Yes
G	3	8	17	11	20	9	No
H	5	15	15	20	20	0	Yes
I	4	15	16	19	20	1	No
J	7	20	20	27	27	0	Yes

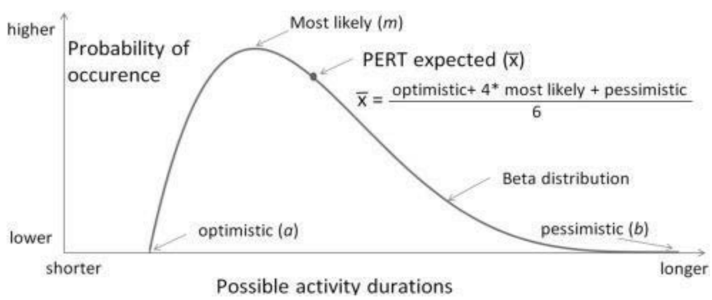
Probabilistic Activity Times

- Activity time estimates usually cannot be made with certainty.
- PERT used for probabilistic activity times.
- In PERT, three time estimates are used: most likely time (m), the optimistic time (a), and the pessimistic time (b); using Beta Distribution.
- These provide an estimate of the mean and variance of a beta distribution:

Variance : $v = \left(\frac{b - a}{6}\right)^2$

Mean (expected time): $t = \frac{a + 4m + b}{6}$

Figure 9: Probabilistic Time Estimates



Example

Mwanza Textile Company has decided to install a new computerized order processing system that will link the company with customers and suppliers online. In the past, orders for the cloth the company produces were processed manually, which contributed to delays in delivering orders and resulted in lost sales. The company wants to know how long it will take to install the new system. We will briefly describe the activities and the network for the installation of the new order processing system.

Mwanza Textile Company – Activities are as follows: The network begins with three concurrent activities: The new computer equipment is installed (activity 1); the computerized order processing system is developed (activity 2); and people are recruited to operate the system (activity 3). Once people are hired, they are trained for the job (activity 6), and other personnel in the company, such as marketing, accounting, and production personnel, are introduced to the new system (activity 7). Once the system is developed (activity 2), it is tested manually to make sure that it is logical (activity 5).

Following activity 1, the new equipment is tested, and any necessary modifications are made (activity 4), and the newly trained personnel begin training on the computerized system (activity 8). Also, node 9 begins the testing of the system on the computer to check for errors (activity 9). The final activities include a trial run and changeover to the system (activity 11) and final debugging of the computer system (activity 10).

Table 5: Mwanza Textile Company – Activities

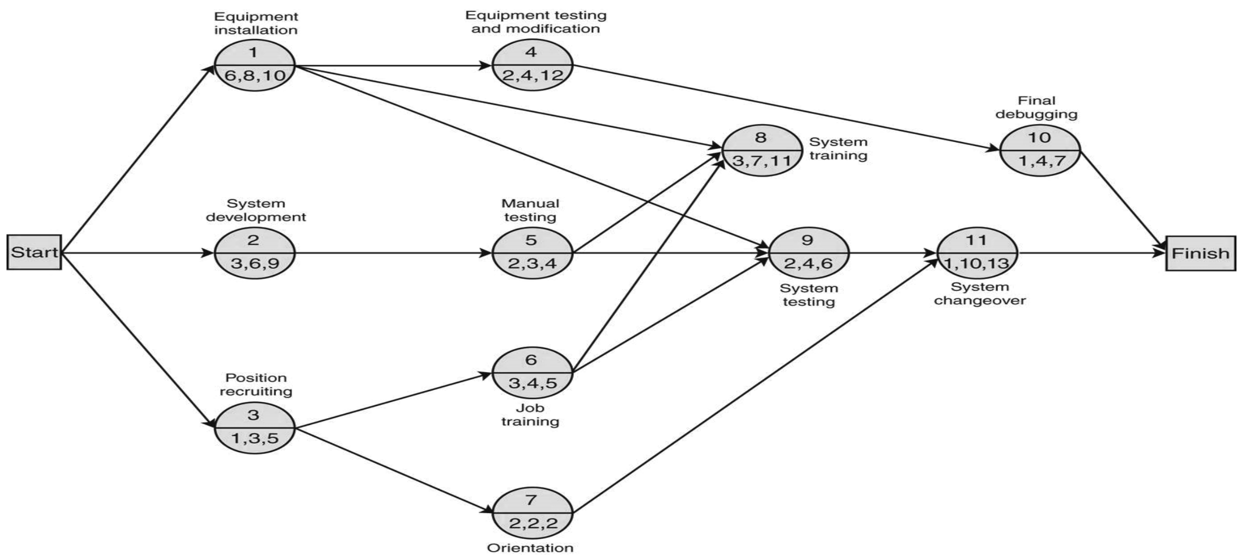
Task	a	m	b	Preceding Tasks		
Task 1	6	8	10			
Task 2	3	6	9			
Task 3	1	3	5			
Task 4	2	4	12	Task 1		
Task 5	2	3	4	Task 2		
Task 6	3	4	5	Task 3		
Task 7	2	2	2	Task 3		
Task 8	3	7	11	Task 1	Task 5	Task 6
Task 9	2	4	6	Task 1	Task 5	Task 6
Task 10	1	4	7	Task 4		
Task 11	1	10	13	Task 7	Task 8	Task 9

Probabilistic Activity Times

Table 6: Mwanza Textile Company

	Time Estimates (weeks)			Time	Mean Variance
Activity	*a*	*m*	*b*	*t*	*v*
1	6	8	10	8	4/9
2	3	6	9	6	1
3	1	3	5	3	4/9
4	2	4	12	5	25/9
5	2	3	4	3	1/9
6	3	4	5	4	1/9
7	2	2	2	2	0
8	3	7	11	7	16/9
9	2	4	6	4	4/9
10	1	4	7	4	1
11	1	10	13	9	4

Figure 10: Mwanza Textile Network for order processing system installation



Expected project time is the sum of the expected times of the critical path activities.

i) Project variance is the sum of the critical path activities' variances
ii) The expected project time is assumed to be normally distributed (based on central limit theorem).
iii) In example, expected project time (tp) and variance (Vp) interpreted as the mean (:) and variance (δ^2) of a normal distribution:

Critical Path Activity	Variance
2	1
5	1/9
8	16/9
11	4
	62/9

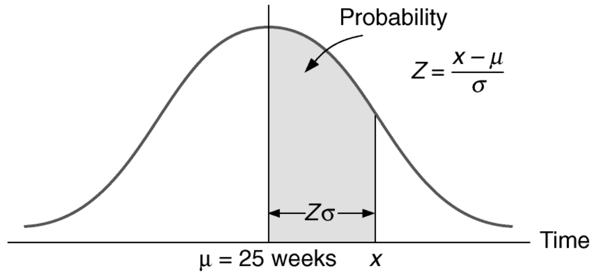
μ **= 25 weeks**

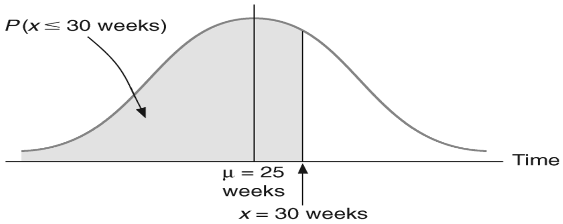
$\sigma^2 = 62/9$

$= 6.9$ **weeks**^{2}

Using the normal distribution, probabilities are determined by computing the number of standard deviations (Z) a value is from the mean. The Z value is used to find the corresponding probability

Figure 11 (a, b): Normal distribution of network duration



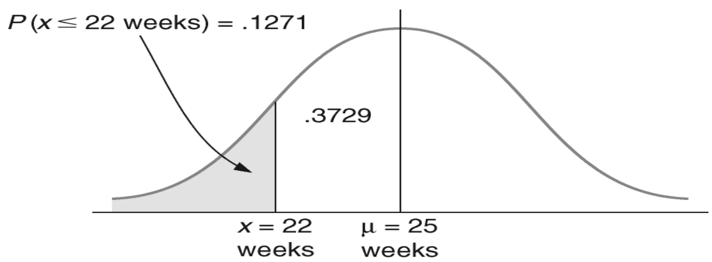


What is the probability that the new order processing system will be ready by 30 weeks?

$$\mu = 25 weeks$$
$$\sigma^2 = 6.9$$
$$\sigma = \sqrt{6.9} = 2.63$$
$$Z = \frac{x - \mu}{\sigma}$$
$$Z = \frac{30 - 25}{2.63} = 1.90$$

Z value of 1.90 corresponds to probability of .4713 in Table A.1, Appendix A. The probability of completing project in 30 weeks or less: (.5000 + .4713) = **.9713.**

What is the Probability that the network will be completed in 22 weeks or less?



A customer will trade elsewhere if the new ordering system is not working within 22 weeks. What is the probability that she will be retained? Z = (22 - 25)/2.63 = -1.14

Z value of 1.14 (ignore negative) corresponds to probability of .3729 in Z Table.
Probability that customer will be retained is **.1271** (.5000-.3729)

Crashing of a Project Activities

Crashing of a project means intentionally reducing the duration of a project by allocating more resources to it. A project can be crashed by crashing its critical activities (because the duration of a project is dependent upon the duration of its critical activities). The use of more workers, better equipment, overtime, etc would generate higher direct costs for individual activities. However, shortening the overall time of the project would also reduce certain fixed and overhead expenses of supervision, as well as indirect costs that vary with the length of the project.

We know that by adding more resources, the duration of an activity can be reduced. If an activity gets completed in ten days with five men working on it, the same activity can be finished in say, six days with ten men (exact mathematical relationships don't work here) working on it. The initial direct cost was 50 man days (5 men x 10 days) and now it is 60 man days (10 men x 6 days). Therefore, the direct cost has increased by 10 men- days.

At the same time, because of the decrease in duration of the activity by four days, the indirect cost (cost of supervision) decreases. Hence, we can conclude that the direct and indirect costs are inversely proportional to each other, i.e. when one increases, the other decreases.

An activity can be crashed by adding more resources only up to a definite limit. Beyond this limit, the duration of the activity does not decrease by adding more resources. This is due to decreasing efficiency of labour, and also increasing confusion due to a large number of resources. In our example, if we increase the number of workers to 15, the same activity can probably be done in four days; but by adding ten more men (so that 20 men work on this activity), the activity time may not decrease further.

The crash time is the shortest time that could be achieved if all effort (at any reasonable cost) were made to reduce the activity time. The limit beyond which the duration of the activity does not decrease by adding any amount of resources is called the **crash time**. It is the shortest possible activity time. The direct cost associated with each crash time is called the **crash cost**.

The **normal time** (10 days in our example) can be defined as the duration of an activity when the minimum possible resources required for its performance are deployed. The lowest expected activity costs are called the **normal costs.**

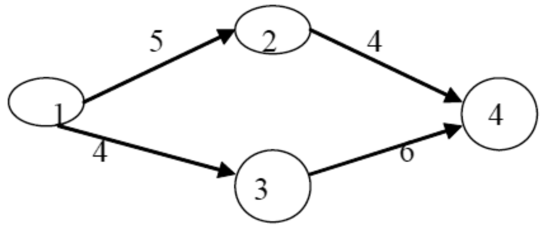
Project direct cost is the direct cost involved in all the activities of the project. **Project indirect cost** is the costs associated with sustaining a project. They include the cost of supervision during the implementation of the project, overheads, facilities, penalty costs and lost incentive payments. The salaries paid to the project manager/supervisor etc. miscellaneous costs due to delays in the project, and rewards to the project team members for its early completion are indirect costs. Project indirect cost is dependent upon other length of duration of the project. A project having a longer duration will have a higher indirect cost (due to supervision required for longer duration).

In any projects, there is a direct relationship between the time taken to complete an activity and the project cost. On one hand, it costs money to expedite a project. Costs associated with expediting a project are called **activity direct costs**, and are different from project direct costs. Some examples of activity direct costs are – hiring more workers, buying or leasing additional equipment, drawing on additional support facilities etc.
If activity direct costs will rise, project indirect costs will fall. Therefore, in a real situation, we need to have a time cost trade-off, this means the sum of activity direct costs and project indirect costs must be minimum.

During the process of crashing of a project, the critical path may get changed. At some stage of crashing, there may even be two or more critical paths (having the same duration) simultaneously. In such situations, one activity is chosen from each of the critical paths and these activities are crashed by unit time to reduce the duration of the project by unit time.

Time-cost models search for the optimum reductions in time. *We seek to shorten the length of a project to the point where the savings in indirect project costs is offset by the increased direct expenses incurred in the individual activities***.**

Example
A network has four activities with expected times as shown. The minimum feasible times and cost per day to gain reductions in the activity times are shown. If fixed project costs are TZS 90 per day, what is the lowest cost time schedule? (***Note: all cots are in, Thousands***)
Figure 12: Network Diagram for Project H.



Activity	Minimum Time in days	Direct Costs of Time Reduction (Tshs)
1-2	2	40 (each day)
1-3	2	35 (first day, 80(second day)
2-4	4	None possible
3-4	3	45 (first day), 110 (other days)

Solution:
First we must determine the critical path and critical path time cost.

Path	Path Times	Total Project Cost (TZS)
1-2, 2-4	5 + 4 = 9	10 days x TShs90/day = TShs 900
1-3, 3-4	4 + 6 = 10	

For ease of reference, let us call the paths A and B respectively. Path B that is, 1-3-4 is the critical path with duration 10 days and cost TZS. 900/-

Next, we must select the activity that can reduce critical path time at the least cost. Select activity 1- 3 at TShs 35 per day, which is less than the TZS90 per day fixed cost.

Reduce activity 1–3 to 3 days. Revise the critical path time cost

Revised Path Times	Total Fixed Cost	Savings over Previous Schedule (TZS)
A: 5+ 4 = 9	9 × TZS.90 = TZS 810	TZS 900 – (810 + 35) = TZS.55
B: 3 + 6 = 9		

Both paths are now critical, so we must select an activity on each path. Select activity 1–2 at TZS 40 per day and 3–4 at TZS 45 per day. Reduce activity 1–2 to 4 days
and 3–4 to 5 days. Revise the critical path time and cost

Revised Path Times	Total Fixed Cost	Savings over Previous Schedule (TZS)
A: 4+ 4 = 8	8 x 90 = TZS 720	810 – (720 + 40 + 45) = TZS 5
B: 3 + 5 = 8		

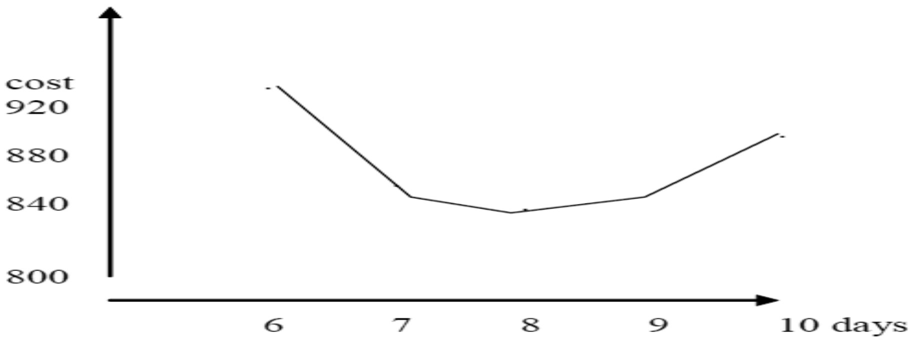
Let us see if we can reduce the time of both paths any further. Activity 1–2 is a good candidate on path A, for it is still at 4 days and can go to 3 for a TZS 40 cost.

But when this cost is combined with the TZS 80 cost for reducing activity 1–3 another day, the sum is greater than TZS 90, so further reduction is not economically justified.

The final step in time – cost analysis is to compare the crash times and the costs associated with them (crash costs). A sufficient number of intermediate schedules are computed such that the total of the direct and indirect (fixed) project costs can be plotted.

Example 6
Graph the total relevant costs for the previous example, and indicate the optimal time – cost trade-off value.

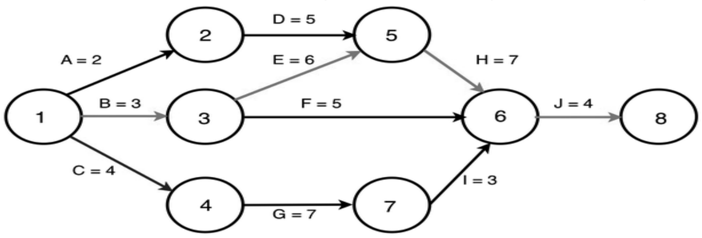
Project Length (days)	Indirect Cost	Activity Reduced	Relevant Direct Cost	Relevant Total Cost
10	900	None	0	900
9	810	1 – 3	35	810 + 35 = 845
8	720	1 – 2 and 3 – 4	35 + 85 = 120	720+40+45+35 = 840
7	630	1 – 2 and 1 – 3	120 + 120 = 240	870
6	540	1 – 2 and 3 – 4	240 + 150 = 390	930



This graph is called the crash – time diagram for completing the project. The lowest total cost is to complete the project in 8 days at a cost of TZS 840,000. However, extending it to 9 days adds only R TZS 5000 to this cost.

Do it Yourself Exercise

1. Determining the Critical Path for a network diagram of a project with Figure X below



Assume all durations are in days.

i) Determine how many paths are on this Network Diagram.
ii) How long is each path?
iii) Which is the critical path?
iv) What is the shortest amount of time needed to complete this project?

INVENTORY PLANNING AND CONTROL

11

Get Through Intro

Material constitutes a significant portion of total costs and hence any increase or decrease has a bearing on profitability. Proper recording (in terms of quantity and value) of all the transactions relating to materials is essential. This includes ordering, receiving and issuing raw materials as well as identifying balances after each transaction together with close monitoring of the physical inventory.

Maintaining appropriate levels of inventory is an important function of inventory management. If you maintain inventory which is too large, excessive capital is tied up (your cash is effectively sitting as inventory in your warehouse). On the other hand, if the inventory is too low, you face the risk of stock outs and production stoppage. Therefore, you have to strike a balance. You also have to determine the optimum quantity to be ordered at a time, considering the stock levels and the consumption rate.

In the future, you may be required to assist management in setting up a good stock control system or improve the existing one, and the concepts discussed in this Study Guide will be prove to be immensely useful.

Learning Outcomes

a) Explain the meaning and objective of material management.
b) Identify the costs involved in an inventory model and determine EOQ for a simple inventory model.
c) Determine EOQ for inventory and production models, lot size, planned shortage and quantity discount models.
d) Explain and apply appropriate methods for establishing Economic Order Quantity and Reorder Levels.

1. Explain the meaning and objective of materials management.
   Identify the costs involved in an inventory model and determine EOQ for a simple inventory model.

[Learning Outcomes a and b]

Materials management is the process of planning, organising, and controlling all those activities concerned with the flow of materials into an organisation. The scope of materials management is very large and includes variety of activities which includes material planning and control, production planning, planning purchases, inventory control etc. All these activities contribute to the undisturbed and smooth running of the operations and minimising the various costs in the organisation.

1.1 Objective of materials management

(a) Obtaining the right quality of raw material: the material needs to be matched to the product specification as it is ultimately reflected in the quality of the finished product.

Compromising quality in return for a lower price is not desirable. The only consideration for procuring the raw material should be to have the right quality and quantity of the item as per the specification, and at the right price.

(b) Not having stock-outs or zero-inventory: this position should be avoided so that production does not come to a halt suddenly, due to shortage or non-availability of the required material. Although stock-out costs are not recorded in the books of accounts and are notional costs, they affect a company in other ways

– e.g. it cannot sell the volume it could have, if the stock-out had not occurred. You also lose the goodwill of a customer if you cannot supply promised goods.

(c) Not holding excess inventory: this should be avoided since this will unnecessarily cause money to be locked up in inventory, and moreover could lead to the deterioration of the unused inventory and cause its subsequent obsolescence.

(d) Ensuring minimum wastage of raw materials: during the storing process or manufacturing wastage should be minimised. Correctly stored materials and efficient use of raw materials in the manufacturing process is essential for this purpose.

Having understood the objectives of material management, let us see the procedures involved in inventory control. Any inventory control system starts from the procurement of raw materials. The ordering of the required material is the first step towards their procurement. In this Study Guide, we will concentrate on procurement and maintenance of the optimal level of inventory. This will ensure that material cost is kept to a minimum, thus ensuring the maximum profitability.

To summarise

The study of inventory models is concerned with two basic questions:

(a) How much should be ordered each time?
(b) At what level should the reordering occur?

The first question relates to determining economic order quantity (EOQ) and can be answered with an analysis of costs of maintaining certain levels of inventory. (Discussed in Learning Outcome 2)

The second question relates to determining the point of ordering inventory and can be answered by determining the re-order point. (Discussed in Learning Outcome 3)

1.2 Costs involved in inventory management

Generally, the cost relating to materials is the largest cost in a typical manufacturing industry. Inventory has to be controlled, managed and checked with adequate care, which obviously adds to the costs. The main reasons for holding inventory are:

(a) To meet future shortages
(b) To hold inventory that will be sufficient to produce goods to meet the expected demand
(c) To take advantage of bulk purchases
(d) To enable the production process to flow smoothly and efficiently
(e) To meet seasonal fluctuations and variations in availability of material

In order to control material inventory, it is essential to procure material in such quantities so as to strike a fine balance between the storage costs and the transport and set-up costs.

However, the cost of holding inventory includes not only the cost of storage but also the cost of money (interest) blocked in the inventory. In this Learning Outcome, you will learn the cost involved in holding raw material.

There are two types of costs which are associated with inventory: the cost of making a purchase and the cost of holding the goods in inventory. These are known as ordering costs and carrying costs respectively.

1. Inventory ordering cost or order placing cost

Ordering costs of inventory include the costs of placing a purchase order, including:

(a) clerical costs of preparing a purchase order
(b) cost of receiving the material
(c) material inspection or testing costs

These are explained in detail below:

(a) Order placing costs

The cost of placing an order involves the cost of order processing, costs of correspondence and communication, advertising costs for inviting tenders, tender evaluation costs etc. These costs are directly proportional to the number of orders to be placed in a year. To save on this cost, the organisation needs to keep the number of orders to a minimum.

Example

The ordering cost per order is Tshs100,000. This includes all the costs of placing the order, order processing, advertising and tender evaluation. The total orders placed in a year are 10; hence, the total ordering cost for the year comes to Tshs1,000,000 (Tshs100,000 x 10).

On the other hand, if the entire annual demand is ordered at one time, the number of orders reduces from 10 to 1 in a year, and then the total ordering cost for the period comes to Tshs100,000 (Tshs100,000 x 1).

There is an annual saving of Tshs900,000.

An increase in the number of orders increases the costs of order processing, advertising for tenders, communication and tender evaluation.

(b) Costs of receiving the material

After the order is placed, the vendor will execute the order by supplying the material. This involves transport costs. The cost of receiving is the cost incurred on the transport of material from the supplier to the factory or depot.

This cost will increase only when the number of vehicles / boxes etc. required for transport increases. The capacity of a carrier is generally fixed and any consignment / batch within this capacity will cost the same for transportation. If the transportation increases directly with the number of purchase orders in a year and vice versa, it becomes relevant to be included in the ordering cost.

Example

In Beta Co, the cost of one container sending from one location to another is Tshs500,000. The capacity of the carrier is 5000 kilograms. This cost of transportation will remain Tshs500,000 for any volume of material between 0-5000 kilograms. It will increase only when the size of the batch increases beyond 5,000 kilograms.

So if Beta Co orders 1,000 kilograms of material in January and another 1,000 kilograms in June, it will incur a cost of Tshs1,000,000 (Tshs500,000 + Tshs500,000) for two trips. If Beta Co had ordered 2,000 kilograms of material in January itself, they would have incurred only Tshs500,000.

(c) Material inspecting or testing cost

The inspection cost of material affects the total ordering costs only when inspection is taken up on a sample basis, where only a sample of the material is inspected from each purchase lot. In this case, if the number of batches ordered in a year increases, the inspection or testing will also increase. This will increase the inspection cost and consequently the ordering cost of the material.

Example

If the inspection or testing cost per lot is Tshs50,000, and the total number of lots to be tested is 10, then the total inspection or testing cost for the year becomes Tshs500,000.

If however, only 5 lots were bought, the inspection cost would be halved. Again if the number of inspections or testing increases from 10 to 20 in a year, the total inspection or testing cost for the period becomes Tshs1000,000.

There is an additional cost of Tshs500,000.

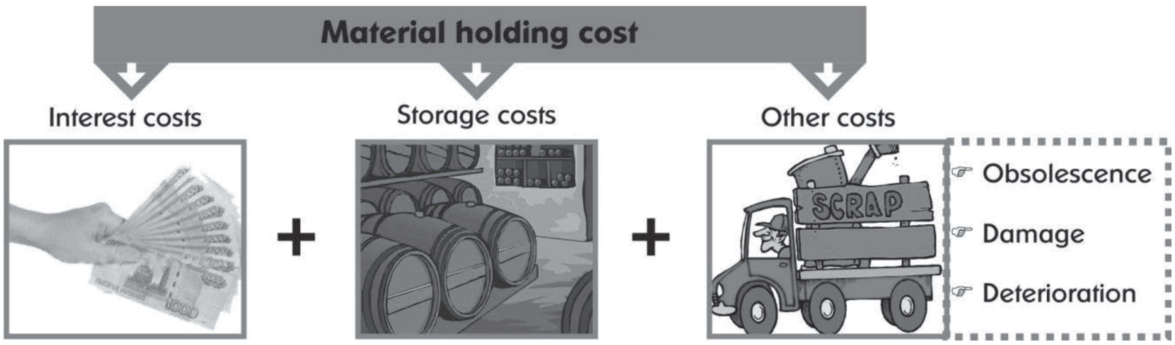
If the entire batch is inspected, the inspection costs do not form part of the ordering cost as it is not dependent on the number of orders. In this case, the inspection cost is considered to be a separate overhead cost.

The above costs of ordering, receiving and inspection together make up the total "ordering cost".

2. Holding (carrying) cost of inventory

The following costs make up the cost of holding inventory:

Diagram 1: Cost of holding materials



(a) Interest costs

Every purchase requires payment to a supplier. This is the investment made in inventory. If the amount required for purchase is borrowed from a bank or from a lender then the interest payable is the cost of investment in inventory.

Example

Suppose a working capital loan of Tshs100,000,000 is taken at 8% from the bank for purchasing raw materials, the interest of Tshs8,000,000 (Tshs100,000,000 x 8%) paid on this loan is the cost of investment in inventory.

(b) Storage cost

Storage costs mainly include costs relating to renting the premises, insurance for the premises and for the inventory. The insurance cost is included in the holding costs only if it varies with a variation in the size of the batch.

Normally the rent cost is fixed for an area irrespective of the volume of material stored in it. If the material ordered exceeds the capacity of the storage space, additional space needs to be hired. This additional space will often cost more.

Example

The rent paid for 500 sq. ft. of warehouse space that can accommodate 1,000 kilograms of material is Tshs500,000. This is the storage cost.

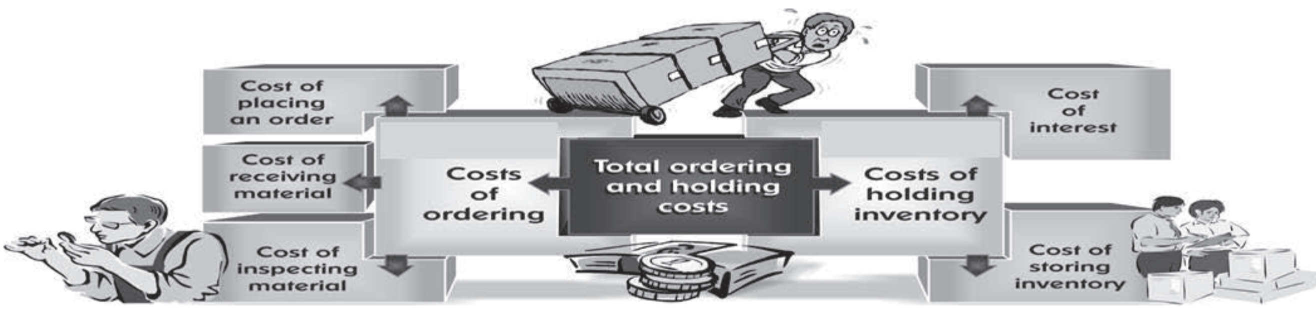
The inventory in store needs to be insured against any mishaps such as natural calamities, fire, floods etc. The cost of insurance depends upon the value of inventory insured. Normally, the entire inventory is covered by insurance. This cost increases when the volume of inventory in storage increases.
If the sizes of the batches ordered are large, then automatically the volume of inventory in storage will increase and hence the cost of insurance will also increase.

Example

The insurance cost is 1% of the value of inventory insured. The insurance cost of Tshs50,000,000 worth of inventory comes to Tshs500,000 (Tshs50,000,000 x 1%). If the company decides to increase the holding of inventory to Tshs100,000,000, then the insurance cost will increase to Tshs1,000,000 (Tshs100,000,000 x 1%).

The cost of ordering and holding inventory can be distinguished as follows:

Diagram 2: Ordering and holding costs of inventory



(c) Other costs

Other costs include risk of obsolescence, deterioration and theft. When material and components become outdated and/or useless, the existing inventory should be thrown away and its cost must be written off in the statement of profit or loss. In the case of theft, the loss arising from theft should be written off in the statement of profit or loss.

1.3 Buffer inventory

There is a possibility that unpredictable events such as poor quality of supplier's product, or poor delivery may disturb the smooth functioning of the inventory system. Buffer inventory is the inventory held on hand that is over and above the currently needed inventory. It is the minimum amount of inventory required to be maintained in order to avoid uncertainties of supply and demand. It is also called the safety stock. The cost of holding buffer inventory is also considered a part of the total holding and ordering cost of inventory.

All the above costs together make up the total inventory holding cost.

Test Yourself 1

The inventory holding cost is the total cost of the investment in inventory, and includes:

A The salary of the watchman who guards the warehouse.
B The salary of the factory supervisor.
C The manager's salary.
D Interest cost for capital which is borrowed for the purchase of inventory.

The formulation of optimum inventory with reference to the above costs can be well explained through an Economic Order Quantity (EOQ) model. EOQ model is discussed in detail in Learning Outcome 2.

2. Determine EOQ for inventory and production models, lot size, planned shortage and quantity discount models.

[Learning Outcome c]

2.1 EOQ for inventory and production model

When an organisation follows a system of ordering fixed amounts of inventory, the order size is vital. This is because the size of the order affects the ordering as well as the carrying costs, and therefore the cash flow of the company too.

When the order size is large, the number of orders required to be placed in a year will reduce, and hence the ordering cost will also reduce. However, in this case the holding costs will increase, as the quantities to be held at the same time increase.

When the order size decreases, the holding costs will decrease but the number of orders per annum increases, thereby increasing the ordering costs.

In this situation, it is vital to arrive at an optimal order size that minimises the total ordering and holding costs of inventory.

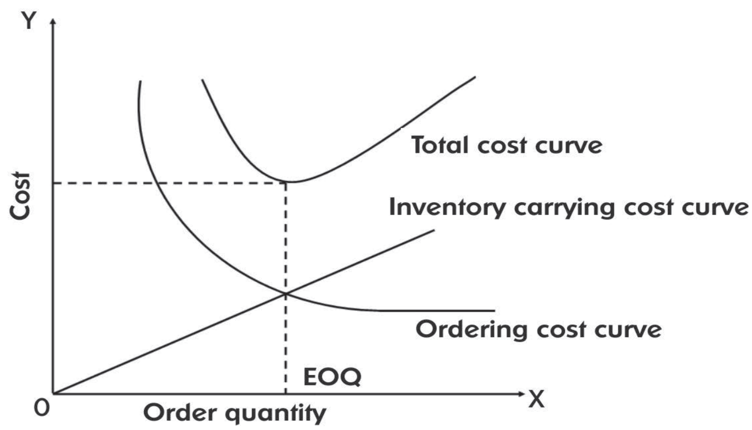
Diagram 3: Relationship of carrying costs and ordering costs

Definition

The optimal re-order quantity or the economic order quantity (EOQ) is a size of order for which the total of the ordering and carrying costs is at the lowest possible.

Economic order quantity can be understood with the help of a graph:

Ordering quantity is measured on the X - axis and cost on the Y - axis.



The graph shows the total cost curve, the carrying (holding) cost curve and the ordering cost curve. In the above graph, the total cost starts at a high level when the ordering cost is initially high and the holding cost is low. When the ordering cost is low and the holding cost is high the total cost is on a rising trend. This happens when one cost is minimised the other cost rises, and vice versa.

However, there is a point where the total cost curve is at its minimum level. If a straight line is drawn from this point passing through the ordering cost line as well as the holding cost line, it passes through the intersection point of these two lines. It implies that the total cost is at its minimum when ordering and holding costs are equal. This is an important observation based on which the formula for the Economic Ordering Quantity (EOQ) can be derived.

When deriving the EOQ, there are some assumptions underlying it without which the formula would not hold true.

(a) The annual demand is certain and known.
(b) The time required for the receipt of material (known as lead time) ordered is certain.
(c) There is no situation of stock outs.
(d) The entire material ordered is received in a single batch.
(e) The per unit cost of material does not change.
(f) The costs are always known with precision.

$$\boxed{EOQ = \sqrt{\frac{2 \times D \times C_0}{C_h}}}$$

Where,
C_0 = cost of ordering per order / consignment from supplier
C_h = cost of holding per unit of inventory per annum / time period
D = total demand during the period

The underlying data for the calculation of EOQ has to remain the same throughout the period for which the calculations are made.

The average inventory is taken to calculate the annual holding cost as one does not hold units equal to the optimal re-order quantity at all times. We might hold inventories greater than this quantity or less than this quantity. In order to calculate the annual holding cost we multiply the holding cost per unit per annum with the average level of inventory. This average level is arrived at by dividing the EOQ by 2.

Example

Calculate the economic order quantity from the following information. Also state the number of orders to be placed in a year and explain briefly the amounts you have calculated.

Consumption of material per annum	10,000 Kilograms
Order placing cost per order	Tshs50,000
Cost per kilogram of raw material	Tshs2,000
Storage cost	8% of material cost

Answer

$$EOQ = \sqrt{\frac{2DC_o}{C_h}}$$

$$\sqrt{\frac{2 \times 10{,}000 \times \text{Tshs}50{,}000}{\text{Tshs}2{,}000 \times 8\%}}$$

$$= \sqrt{\frac{1{,}000{,}000}{0.16}}$$

$$= \sqrt{62{,}50{,}000}$$

= 2,500 kilograms

No. of orders to be placed in a year = Consumption of material per annum / EOQ
= 10,000 kilograms/2,500 kilograms
= 4 orders per year

4 orders need to be placed per year with an order size of 2,500 kilograms to keep the ordering and the holding costs at the minimum level.

The above example shows how to calculate the optimal re-order quantities. Another example will be helpful for further understanding.

Example

Tubes Plc manufactures picture tubes for televisions. Details of their operation during 20X3 are as follows:

Normal weekly usage	100 tubes
Ordering cost	Tshs50,000 per order
Inventory holding cost	20% per annum
Cost of tubes	Tshs300,000 per tube

Calculate the optimal re-order quantity.

Answer

The EOQ calculation requires details of the annual use in units, ordering cost per order and the carrying cost per unit per annum. These will be calculated as follows:

Annual use (A) = Weekly usage x number of weeks in a year
= 100 tubes x 52 weeks
= 5,200 tubes

Ordering cost per order (O) = Tshs50,000 per order (given)

The carrying cost is given in the question as a percentage of the cost of tubes.
Cost of one tube = Tshs300,000 (given)

Carrying cost per tube per annum (C) = Tshs300,000 x 20%
= Tshs60,000 per tube per annum

Therefore EOQ = $\sqrt{\dfrac{2DC_o}{C_h}}$

$\sqrt{\dfrac{2 \times 5{,}200 \text{ tubes} \times \text{Tshs}50{,}000}{\text{Tshs}60{,}000}}$

$\sqrt{\dfrac{\text{Tshs}5{,}200{,}000}{\text{Tshs}60{,}000}}$

= $\sqrt{8{,}666.67}$

= 93.09 or 93 tubes approximately

Therefore, the optimal re-order quantity is 93 tubes.

Tip

It was discussed above that the ordering and carrying costs are equal at the EOQ. Let us verify it.

In the above question, the EOQ is calculated as 93 tubes approximately. For the purpose of the calculation we will take the absolute figures without rounding them up. The EOQ will then be 93.09.

Annual ordering cost = $\dfrac{\text{Ordering cost per order x Annual usage}}{\text{EOQ}}$

$\dfrac{(\text{Tshs}50{,}000 \times 5{,}200 \text{ tubes})}{93.09}$

= Tshs2,792,996

Holding cost of tubes per annum = Holding cost of tubes per tube per annum x Average tubes held at all times

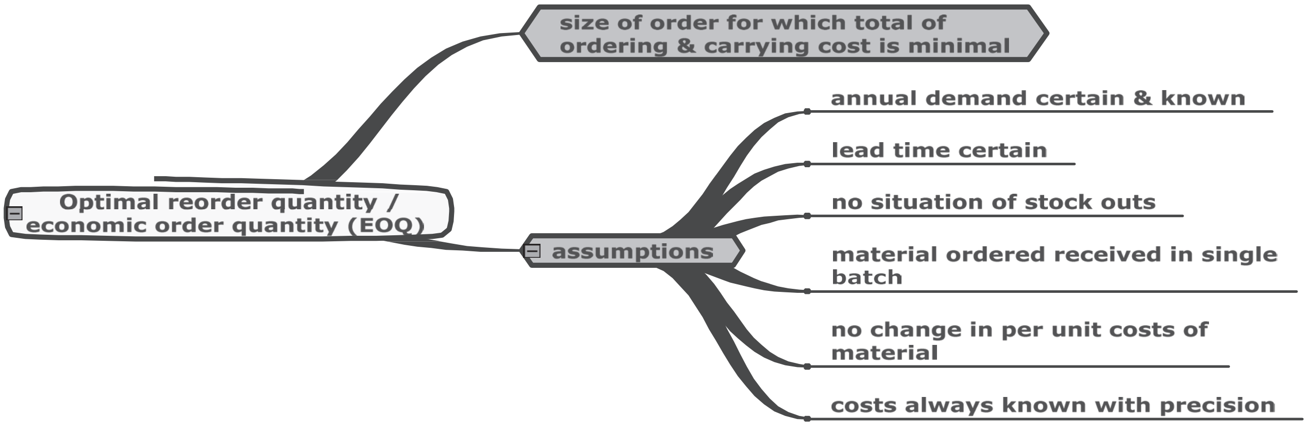
The holding cost per order will therefore be = $\dfrac{\text{EOQ x holding cost per tube per annum}}{2}$

$\dfrac{93.09 \times \text{Tshs}60{,}000}{2}$

= Tshs2,792,996

Hence it can be seen from the above that the ordering and the holding costs are equal at the EOQ.

SUMMARY



Test Yourself 2

Bearings Plc committed to supply 24,000 bearings per annum to Motor Plc. It is estimated that it costs Tshs100 as inventory holding cost per bearing per month, and that the set-up cost per run of bearing manufacture is Tshs324,000.

The optimum run size for Bearings Plc would be:

A 4,600 bearings
B 6,300 bearings
C 3,600 bearings
D 3,000 bearings

2.2 Determining EOQ when discounts are received

The above model of EOQ assumes that the unit price of materials remains the same throughout the year. However there may be situations where the unit price of materials changes when the supplier offers discounts in price. Vendors / suppliers offer different levels of quantity discounts for different volumes of materials ordered. These discounts are also referred to as quantity discounts.

If discounts apply, the formula for EOQ as derived earlier will not hold true. The EOQ will have to be arrived at by computing the total cost (total of the ordering and the holding cost) at various levels of material ordered. After this, the quantity at which the overall costs are the least can be chosen to be the optimal quantity to be ordered.

The steps to be followed for determining the EOQ when discounts are available on the quantity of order are as follows:

Step 1: Calculate the Economic Order Quantity, without considering discounts.

Step 2: If the Economic Order Quantity calculated above is less than the minimum order quantity required to avail the discount, calculate the Total Annual Costs for the EOQ obtained above and for the minimum order size required to avail the discount.

Step 3: Compare the Total Annual Costs for both the quantities i.e. the EOQ and the minimum order size required to avail the discount. Select the quantity with minimum Total Annual Cost.

Step 4: If there is a further discount available for an even larger order size, repeat the same calculations for the higher discount level.

Step 5: The order size with the least Total Annual Cost shall be ordering quantity..
This concept can be best understood with the help of the following example.

Example

A firm is eligible to quantity discounts on its orders of material as follows:

Price per tonne (Tshs('000))	Tonnes
8.0	Less than 500
7.9	500 and less than 1500
7.8	1500 and less than 3000
7.7	3000 and less than 5000
7.6	5,000 and over

(a) The annual demand for the material is 5,000 tonnes.

(b) Inventory holding costs are 15% of material cost per annum.

(c) The delivery cost per order is Tshs7,000.

Required:

Calculate the optimal quantity to order at which costs are kept to a minimum.

Answer

In the given problem we have been provided with the discounts that the firm can avail for different ordering quantities.

In this case the optimal re-order quantity will have to be calculated by computing the total costs at each order size taking into consideration the discounts offered. The case where the total cost is the least will be the optimal re-order quantity for material orders.

Let us see, the calculations for total cost at the ordering quantity of 100 units:

Ordering quantity = 100 units (given in the problem)
Price per tonne = Tshs 8,000 (given in the problem)
Purchasing cost of 5,000 tonnes = 5,000 x Tshs8,000 = Tshs4,000,000
No of orders to be placed = Annual consumption / ordering quantity = 5000/100 = 7 orders
Inventory holding cost = 15% of purchase price (given in the problem)
Therefore, Inventory holding cost = (Average Inventory) x 15% of purchase price
= (Ordering quantity/2) x 15% of purchase
= (100/2) x 15% x Tshs8,000
= Tshs60,000
Total cost = Purchase price + ordering cost + Inventory holding cost
= Tshs(40,000,000 + 350,000 + 60,000)
= Tshs40,460,000

Similarly, the total cost for all the ordering quantities is calculated in the table given below:

Ordering quantity (tonnes)	Price per tonne (Tshs ('000)	Purchasing cost of 5,000 tonnes	No of orders to be placed during the year	Ordering cost (Tshs'000)	Inventory holding cost (Tshs'000)	Total cost (Tshs ('000)
		5,000 T x price per tonne	(5,000 T/ ordering quantity)	(No. of orders to be placed x 7)	(Ordering quantity / 2) x purchase price per ton x 15%	Lowest cost
100	8.00	40,000.00	50.00	350.00	60.00	40,410.00
500	7.90	39,500.00	10.00	70.00	296.25	39,816.25
1,500	7.80	39,000.00	3.33	23.33	877.50	39,900.83
3,000	7.70	38,500.00	1.67	11.67	1,732.50	40,244.17
5,000	7.60	38,000.00	1.00	7.00	2,850.00	40,857.00

From the above table we can see that the total cost is the minimum at the ordering quantity of 500 units.

Therefore, the optimal ordering quantity is 500 units.

Note : The ordering quantity is divided by 2 to obtain the average inventory in order to calculate the annual holding cost as one does not hold units equal to the optimal re-order quantity at all times.

Example

The following is the information of ABC Ltd regarding the raw material it consumes in one of its process:

Monthly consumption	150 units
Purchase price	Tshs10,000 per unit
Ordering costs	Tshs50,000
Holding costs	20%
Discount from supplier if order is minimum 2,000 units	5%

Required:

What will be the most economical order size if 5% discount can be received from the supplier for a minimum order of 2,000 units?

Answer

In the given problem, we see that the ABC Ltd shall receive a discount of 5% on the purchase price for a minimum order of 2,000 units. But the annual demand of ABC Ltd is 1,800 units (150 units x 12)

Here, we need determine the most economical ordering quantity for ABC Ltd. And whether ABC Ltd. should order 2,000 units to avail the discount?

This can be done with the help of the following steps.

Step 1: Calculate the Economic Order Quantity, without considering discounts

As per the given problem we need to calculate the EOQ for 1,800 units (Annual demand)

D = annual demand = 150units x 12 months = 1,800 units
C_o = ordering costs = Tshs50,000 (given)
C_h = holding costs = Tshs10,000 x 20% = Tshs2,000

$$EOQ = \sqrt{\frac{2DC_o}{C_h}}$$

$$= \sqrt{\frac{2 \times 1{,}800 \times \text{Tshs}50{,}000}{\text{Tshs}2{,}000}}$$

= 300 units

Step 2: If the Economic Order Quantity calculated above is less than the minimum order quantity required to avail the discount, calculate the Total Annual Costs for the EOQ obtained above and for the minimum order size required to avail the discount.

As per the given problem Economic Order Quantity of 300 units is less than the minimum order quantity required to avail the discount i.e. 2,000 units.

The Total Annual Cost for both the ordering quantities i.e. 1300 units and 2,000 units is calculated in the table given below:

	300 Units (EOQ)	2,000 units (Minimum order size to avail discount)
Purchase cost (W1)	18,000,000	19,000,000
Holding costs (W2)	300,000	2,000,000
Ordering costs (W3)	300,000	4,500
Total annual costs	18,600,000	21,045,000
Cost per unit	10,333.33	10,522.5

Continued on the next page

Step 3: Compare the Total Annual Costs for both the quantities i.e. the EOQ and the minimum order size required to avail the discount. Select the quantity with least Total Annual Cost.

From the table we can see that the Total Annual Cost for 300 units is the minimum.

Step 4: If there is a further discount available for an even larger order size, repeat the same calculations for the higher discount level.

In the given situation there are no further discounts available on any other quantity therefore, no further calculation is required.

Step 5: The order size with the least Total Annual Cost shall be Economic Order Quantity.

As the total annual cost of the order size of 2,000 units is higher, ABC should opt for placing an order of 300 units at a time.

Alternatively, total annual cost can be calculated using the following formula:

$$\text{Total annual cost} = \boxed{P + \left(C_o \times \frac{D}{Q}\right) + \left(C_h \times \frac{Q}{2}\right)}$$

Where,

P = purchase cost
D = demand per annum
Q = reorder quantity
C_o = ordering cost
C_h = holding cost

$$\text{Total annual cost} = P + \left(C_o \times \frac{D}{Q}\right) + \left(C_h \times \frac{Q}{2}\right)$$

$$= (\text{Tshs}10{,}000 \times 1{,}800 \text{ units}) + \left(\text{Tshs}50{,}000 \times \frac{1{,}800 \text{ units}}{300 \text{ units}}\right) + \left(\text{Tshs}2{,}000 \times \frac{300 \text{ units}}{2}\right)$$

= Tshs18,600,000

Workings

W1 Purchase cost

Purchase cost = Units x Purchase price per unit

300 Units (EOQ)	2,000 units (Minimum order size to avail discount)
1,800 units x Tshs10,000 = Tshs18,000,000	(2,000 units x Tshs10,000) – 5% discount = Tshs19,000,000

W2 Holding costs

Holding costs = Average inventory x Holding cost per unit
(Average inventory = EOQ/2)

300 Units (EOQ)	2,000 units (Minimum order size to avail discount)
(300/2) x Tshs2,000 = Tshs300,000	(2,000/2) x Tshs2,000 = Tshs2,000,000

W3 Ordering costs

Ordering costs = Number of orders x Ordering costs per order

300 Units (EOQ)	2,000 units (Minimum order size to avail discount)
(Number of orders = Annual demand/EOQ)	(Number of orders = Units required for discount /Order quantity)
(1,800/300) x Tshs50,000 = Tshs300,000	(1,800/2,000) x Tshs50,000 = Tshs45,000

Test Yourself 3

Star Plc manufactures ice-cream. The following is a schedule showing discounts offered by the supplier of a special flavouring agent.

Quantity (litres)	Price per litre in Tshs ('000)
150–250	25
250-500	24
500-750	23
750-1,000	20
1,000 and above	18

The annual demand is 1000 litres. The ordering cost per order is Tshs150,000 and the carrying cost per annum is 20% of the material cost.

Required:

Calculate the order size at which the ordering and the holding cost is the minimum.

2.3 Determine EOQ for production model or the small lot sizes model

In case of a manufacturer who produces the inventory rather purchasing it determination of the optimal production lot size is a very vital decision to be made by the management. In this case the previous model of EOQ will not hold true for the calculation of the production lot size. Here the inventory required for production of the main product is not received in lump sum as it is not purchased but gradually produced simultaneously with the main product. Here the production rate of the inventory needs to be higher than the consumption or demand rate.

In this model the optimal re-order quantity is known as the Economic Batch Quantity. The Economic Batch Quantity is the production lot size of the inventory that is required to be produced by the manufacturer so that the production of the main product continues smoothly and the holding costs are also low.

In this model the ordering cost is replaced by the set up cost. Setup cost is the cost incurred to prepare a machine or process for manufacturing an order.

This same model is used where a manufacturer orders raw materials rather producing it but the supplier supplies the inventory in small sizes gradually. When this model is used for calculating optimal order size for inventory purchased rather than produced, the set up cost shall be replaced by the normal ordering cost.

This model runs on the basis of a few assumptions given below:

(a) Demand occurs at a constant rate of D items per year or d items per day.
(b) Production rate is P items per year or p items per day
(c) Set-up cost i.e. C_o per run.
(d) Holding cost i.e. C_h per item in inventory per year.
(e) Purchase cost per unit is constant (no quantity discount).
(f) Set-up time (lead time) is constant.
(g) Planned shortages are not permitted.

The formula for the optimal production lot-size or the Economic Batch Quantity under this method is given below

$$EBQ = = \sqrt{\frac{2C_oD}{C_h\,(1 - D/R)}}$$

Where D – Annual demand or requirement of material C_0 – Set up cost per order C_h – Carrying cost per unit per annum R – Rate of production per time period

Example

Natural Plc. produces herbal soaps. The details of the annual demand of the soap and the requirement of the soap and the set up and holding costs are given below.

Annual requirement of the soap = 3600 pieces
Rate of demand = 15 soaps / shift
Number of working shifts in a year = 480 shifts
Cost of the cake mix = Tshs15,000 per piece
Inventory holding cost per annum = 18% of the value of the raw material
Set up cost per order = Tshs60,000

Required:

Calculate the optimal reorder quantity.

Answer

(a) The annual requirement is given as 3,600 pieces.
(b) Per shift rate of demand is given as 15 pieces per shift and number of shifts are 480.
(c) Therefore the rate of production will be = 480 shifts x 15 pieces per shift = 7,200 pieces.
(d) The holding cost is 18% of the value of the raw material = Tshs15,000 x 18% = Tshs2,700
(e) The set up cost per order is Tshs60,000.

The formula for the optimal reorder quantity is given below

$$EBQ = \sqrt{\frac{2C_oD}{C_h\,(1 - D/R)}}$$

Substituting all the values in the formula, we have:

$$\sqrt{\frac{2 \times 3{,}600 \times \text{Tshs}60{,}000}{\text{Tshs}2{,}700\left(1 - \frac{3{,}600}{7{,}200}\right)}}$$

$$EBQ = \sqrt{\frac{432{,}000}{1.35}} = 565.68 = 566 \text{ pieces}$$

The optimal reorder quantity will be 566 pieces.

Test Yourself 4

Hexter Company is a television manufacturer who also makes its picture tubes in-house, for the televisions. The annual requirement of picture tubes is 3,000 units, and 300 units of picture tubes are made per week. Hexter's annual working weeks are 50. The cost of making one picture tube is Tshs20,000, and each time a batch of picture tube is made, Hexter Company has to spend Tshs500,000 towards setting up of machinery. The cost of holding one unit of picture tube is Tshs2,000.

Required:

Calculate the Economic Batch Quantity (EBQ) to the nearest whole digit.

2.4 Determine EOQ when there is a planned shortage

One of the basic assumptions of the EOQ model was that there can be no situations of stock-outs / shortages. However, to arrive at the EOQ with planned shortages, this assumption needs to be changed. The following assumptions need to be replaced to calculate EOQ with planned shortages.

Planned shortages are allowed in cases where basically the holding cost of the inventory is very high. This usually happens in case of high valued products where the investment in maintaining the inventory for finished goods is a very big cost for the management to cope up with. Therefore, in such situations permitting limited planned shortages makes sense from a managerial perspective.

Here the management believes that no customers shall be lost due to unavailability of the product in the market the customers are ready to wait for the product to become available again for e.g. Apple iphones. Their backorders (Back order costs has been explained in the later part of the LO) are noted and fulfilled immediately when the order quantity arrives to replenish inventory."

In this model the manufacturer has a predetermined backorder quantity (Back order quantity has been explained in the later part of the LO). When the orders reach this quantity, the replenishment order arrives for further production.

This model has been designed to determine the optimal order and backorder quantity for any manufacturing concern.

The variable costs in this model are annual holding costs, backorder costs, and ordering costs.

For the optimal order and backorder quantity combination, the sum of the annual holding and backordering costs equals the annual ordering cost.

Annual holding costs and ordering costs are already explained in Learning Outcome 1.

What is backorder cost?

When the company receives back orders from its customers, it incurs backorder cost which includes the real and perceived costs of the inability of the company to fulfil an order. The costs can include negative customer relations, interest expenses, etc. and are typically represented in financial reports on a per-unit basis.

Backorder costs are important for companies to track, as the relationship between holding costs of inventory and backorder costs will determine whether a company should over- or under-produce. If the carrying cost of inventory is less than backorder costs (this is true in most cases), the company should over-produce and keep an inventory.

Back order quantity

The minimum number of orders that a company needs to book, before it resumes production to replenish the inventory in the market is known as the Back Order Quantity.

Example

Rivera Ltd is in the business of selling mobile phones. Since the inventory holding cost is very high, the company maintains a low invemntory. During a festive season, it offers a 30% discount and receives an unexpected number of orders for 5,000 units. Rivera ltd. only has 4,000 units in its warehouse and will need three weeks to make the balance 1,000 units. This situation typically represents a planned shortage. There are some real costs associated with not having products on hand when customers want them (referred to as back order costs).

Some of the costs are tangible. For instance, Rivera Ltd will have to spend a lot of money on expedited shipping to get parts from its suppliers faster in order to fill the order. Then, it may have to pay its laborers a lot of money in overtime in order to hasten production, and when it finally has the 1,000 units ready to ship, it might have to ship overnight those items to customers who are anxious about having the phones in time for the festival. Further, the company's customer service department will likely have some long days and nights which shall result into overtime payment.

Here, the back order quantity is the number of orders that Rivera Ltd. needs to book before starting back production.

The formula for the optimal reorder quantity under this method is given as:

$$\text{EOQ (Q)} = \sqrt{\frac{2DC_o}{C_h}} \times \sqrt{\frac{C_h + C_b}{C_b}}$$

The maximum number of backorders that can be calculated as:

$$S = Q \times \left(\frac{C_h}{C_h + C_b}\right)$$

Backorder costs are the costs associated with being out of stock when an item is demanded (including lost goodwill)

Where D – Annual demand or requirement of material
C_0 – Set up cost per order
C_h – Carrying cost per unit per annum
C_b – Backorder costs per unit per annum
D – Rate of production per time period
S – Back order cost

Example

Global manufactures LED TVs for which the assumptions of the inventory model with shortages are valid. Demand for the product is 2,000 units per year. The inventory holding cost rate is 20% per year. The cost of TV to Global is Tshs50,000. The ordering cost is Tshs25,000 per order. The annual shortage cost is estimated to be Tshs30,000 per unit per year. Global operates 200 days per year.

Required:

Calculate the optimal order quantity for Global and the maximum number of backorders the company can receive.

Answer

The optimal order quantity is calculated as follows:

$$EOQ\ (Q) = \sqrt{\frac{2DC_o}{C_h}} \times \sqrt{\frac{C_h + C_b}{C_b}}$$

Given information is

D – 2000 units
C_0 – Tshs25,000
C_h – 20% of the cost; therefore, Tshs50,000 x 20/100 = Tshs10,000
C_b – Tshs30,000

EOQ (Q) =

$$= \sqrt{\frac{2 \times 2{,}000 \text{ units} \times \text{Tshs}25{,}000}{\text{Tshs}10{,}000}} \times \sqrt{\frac{(\text{Tshs}10{,}000 + \text{Tshs}30{,}000)}{\text{Tshs}30{,}000}}$$

= 115.47 units

The maximum number of backorders can be calculated as follows:

$$S = Q \times \left(\frac{C_h}{C_h + C_b}\right)$$

$$= 115.47 \times \frac{\text{Tshs}10{,}000}{(\text{Tshs}10{,}000 + \text{Tshs}30{,}000)}$$

= 28.87 or 29 units

Test Yourself 5

For the inventory model with planned shortages, which of the following holds true at the optimal order quantity?

A Annual holding cost = annual ordering cost.
B Annual holding cost = annual backordering cost.
C Annual ordering cost = annual holding cost + annual backordering cost.
D Annual ordering cost = annual holding cost – annual backordering cost.

3. Explain and apply appropriate methods for establishing economic order quantity and reorder levels.

[Learning Outcome d]

The Institute of Cost and Management Accountants of England and Wales, defines perpetual inventory as "A system of records maintained by the controlling department, which reflects the physical movement of stocks and their current balances." Setting pre-determined inventory levels is an integral part of the perpetual inventory system, and therefore inventory is regularly reordered once levels go down to the pre-determined level. These levels are set as per the distinct requirements of an organisation. Consequently, defining the levels of inventory differs from organisation to organisation. In this Study Guide we have adopted one of the most widely used models.

3.1 Reorder level

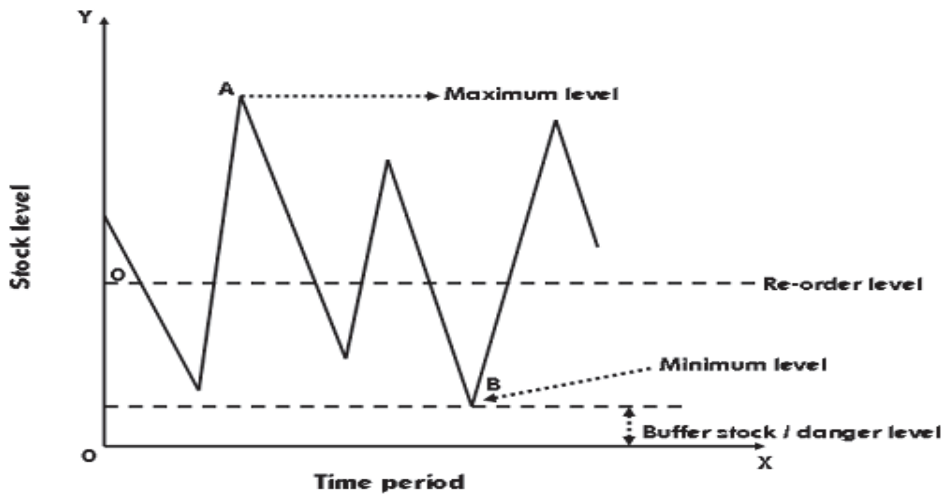
Definition

Re-order level is that predetermined quantity of inventory that when reached initiates a reorder in a perpetual inventory control system.

The calculation of the re-order level is based on certain assumptions:

(a) The material requirements are pre-defined.
(b) In a time interval, the rate of issue of the material is fixed.

A graphical representation of the inventory levels and the time intervals in which the materials need to be re-ordered will explain the situations where the re-order levels apply.



The X-axis measures the time period and the Y-axis measures the level of inventory. The lines depicting the levels of inventory are straight line slopes indicating that the inventory levels go down gradually at a fixed rate. Re-order levels can be calculated only when the demand rate for material is fixed, as given in the assumptions.

As seen in the graph, point A shows that the maximum inventory level, point O, is the re-order level, and point B is the minimum inventory level. The lead time starts from the time when the inventory reaches the re-order level and ends when it reaches the minimum level of inventory.

The re-order level is the level when a fresh order for inventory should be placed for replenishment. By doing so, inventory gets replenished before it reaches the minimum level. We will study the minimum level at a later stage.

Formula for reorder level:

Reorder level= Maximum usage x Maximum lead time

The above formula has been devised so as to suggest that reorder should be given at a level of inventory that would be sufficient enough to avoid a stock-out situation. Even if replenishment of inventory takes maximum possible time and the inventory has been issued to production at the maximum rate during the time period, stock-out will not occur.

OR

Reorder level= Minimum level + (Average usage x Average lead time)

Example

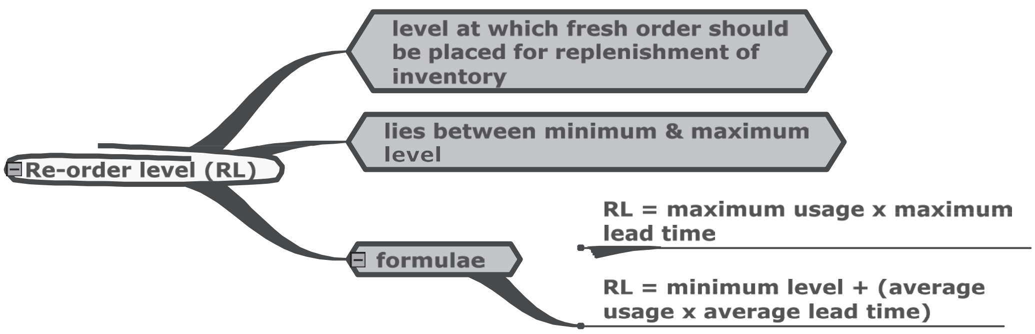
Thames Plc manufactures floating tubes for swimming. They require plastic for manufacturing these tubes. A proper material management system is in place at Thames Plc. The maximum usage of plastic sheets is 75 sheets per week, and the re-order period is 4 to 6 weeks. Calculate the re-order level.

Answer

Maximum usage = 75 sheets of plastic
Maximum lead time = 6 weeks
Using the formula:
Re-order level = Maximum usage x Maximum lead-time
= 75 sheets x 6 weeks
= 450 sheets

Arithmetical calculation of the reorder level requires other figures e.g. maximum reorder period, maximum usage of inventory in units, minimum level of inventory, average rate of consumption of inventory and average lead time. So we will need to understand the meaning of these other terminologies.

SUMMARY



3.2 Minimum level

Definition

Minimum level of inventory is a predetermined quantity of inventory after which any issues of material are made from the buffer inventory if the usage rate is above average.

The above level of inventory is a level where the management needs to take care that the level does not reach a stock-out stage.

The formula used for this calculation is as follows:

Minimum inventory level = Reorder level– (Average usage x Average lead time)

This minimum level indicates the lowest levels of inventory balance, which must be maintained at all times, so that production is not adversely affected due to non-availability of inventory. Any fall in the level of inventory below this point will hamper production, and hence profits.

$$\text{Average delivery period (lead-time) for each item} = \frac{\text{Maximum period} + \text{Minimum period}}{2}$$

Example

If in the above example of Thames Plc, the normal / average usage of material is given as 50 sheets of plastic per week, the minimum level of inventory can be calculated as follows.

Answer

Minimum level = Reorder level – (Average usage x Average lead-time)

We have the reorder level value but the average lead time needs to be computed. The delivery time is given as 4 to 6 weeks. Thus the minimum period is 4 weeks, whereas the maximum period is 6 weeks.

$$\text{Average lead time} = \frac{\text{Maximum period} + \text{Minimum period}}{2}$$

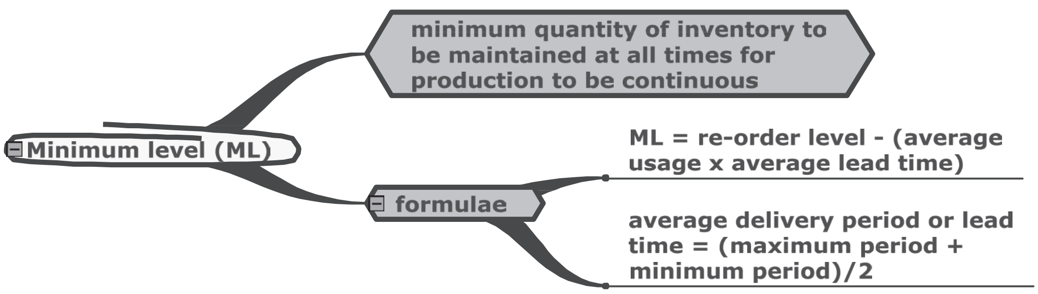
$$= \frac{(4 + 6)}{2} = 5 \text{ weeks}$$

Therefore,

Minimum inventory level = Reorder level – (Average usage x Average lead time)
= 450 – (50 x 5)
= 200 sheets

This indicates that Thames Plc needs to keep a minimum inventory of 200 sheets to avoid stock-outs / problems with production. By the time these 200 sheets are used, the new inventory should be received.

SUMMARY



3.3 Maximum Level

Definition

Maximum level is the level that indicates the maximum quantity of inventory held at any time.

Once this level is reached, it is an indication that the inventory will soon reach a level where it will be overstocked, and cash is unnecessarily tied up in inventory and warehousing.

This is the maximum capacity of the inventory storage. Important factors to consider for the calculation of the maximum level are:

1. Information about its reorder level.
2. Maximum rate of consumption of material and maximum delivery period.
3. Minimum consumption and minimum delivery period for each inventory.
4. Economic order quantity (EOQ).
5. Availability of funds, storage space, nature of items and their price.
6. In case of imported material, due to its irregular supply, the maximum level should be high.

The mathematical formula used for its determination is as follows:

Maximum inventory level = Reorder level + Reorder quantity – (Minimum usage x Minimum lead time)

Example

Stride Plc received a contract to supply 1000 tennis balls to one of London's largest tennis training centres. It needs to know what maximum level of inventory it should hold, so as not to overstock any of the material. The reorder quantity for Stride Plc is 186 units. The delivery period of the material is 2 to 3 weeks. The minimum usage of material is given as 25 units and the maximum usage is 60 units.

Answer

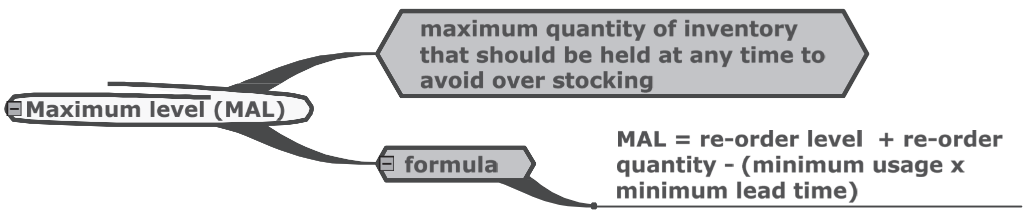
For the maximum level of inventory calculation, we need to know the reorder level.

Reorder level = (Maximum usage x Maximum lead time)
= 60 x 3
= 180 units

Maximum level = Reorder level + Reorder quantity – (Minimum usage x Minimum lead time)
= 180 +186 – (25 x 2)
= 316 units

Stride Plc should hold, at the most, 316 units of inventory in order to avoid overstocking the material and locking the money up in the inventory.

SUMMARY



3.4 Average level

As the name suggests, average inventory means the average of the minimum and maximum levels of inventory.

The average level of inventory is determined by using the following formula:

$$\text{Average inventory level} = \frac{\text{Maximum level} + \text{Minimum level}}{2}$$

OR

Average inventory level = Minimum level +1/2 Reorder quantity

Tip

Remember answer derived using the above two formulae may slightly differ. Both of the answers are acceptable.

Example

In the above example of Thames Plc, we calculated the minimum level to be 200 sheets. In this case, if the reorder quantity is given as 102 sheets, the average level of inventory will be calculated as:

Average inventory level = Minimum level +1/2 Reorder quantity
= 200 + 1/2 (102)
= 251 sheets.

Thames Plc has an average inventory level of 251 sheets. This will give Thames an idea of the average amount of money locked up in inventory.

3.5 Danger level

Definition

Danger level is when the volume of inventory is at the minimum level, and is a trigger for an immediate action of purchase to avoid any stock-out situation.

The formula for the danger level is given below:

Danger level = Average consumption x Lead time for emergency purchases

At the danger level, the inventory levels are so alarmingly low that immediate or emergency purchases need to be made so as not to break the continuity of production.

Example

Precaution Plc manufactures saline solutions. It has faced huge losses in the past due to material shortages of packing bottles. It has recently received a big order to supply 15,000 saline bottles to Polo hospital. It now wants to develop and maintain an accurate system of inventory.

Its first priority is to know the danger level of inventory before it starts production, so that the production does not come to an unexpected halt in the first stage. The average consumption of inventory is 1,500 bottles per month and the lead time for emergency purchases as given is 4 days.

Required:

Calculate the danger level of inventory.

Answer

Average consumption per day is required for calculating the danger level as the lead time for emergency purchases is given in days.

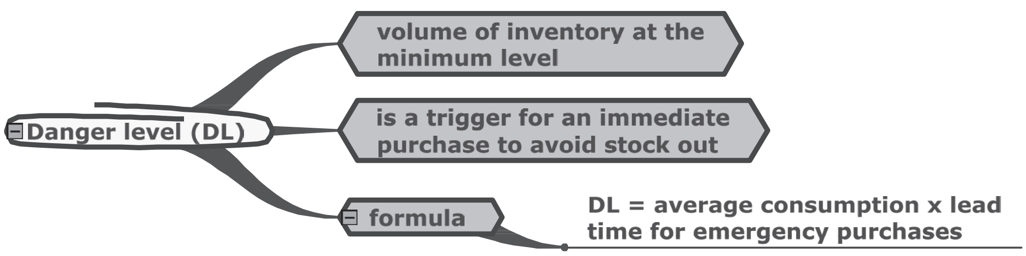
Average consumption = 1,500 bottles per month

Therefore average consumption per day = 1500 / 30 days per month
= 50 bottles per day

Danger level = Average consumption x Lead time for emergency purchases
= 50 bottles x 4 days = 200 bottles

When the level of inventory reaches 200 bottles, Precaution Plc needs to make emergency purchases in order to avoid a stock-out. Care needs to be taken to ensure that the inventory is replenished well before it reaches this level.

SUMMARY



Example

A company uses three raw materials; cotton, nylon and cotton blend for making clothes. The following data applies to it:

Raw material	Usage per unit of product (Kg)	Re-order quantity (Kg)	Price per Kg	Delivery period (in weeks)			Reorder level (Kg)	Minimum level (Kg)
				Min.	Avg.	Max		
Cotton	10	10,000	0.10	1	2	3	8,000	
Nylon	4	5,000	0.30	3	4	5	4,750	
Cotton blend	6	10,000	0.15	2	3	4		2,000

Weekly production varies from 175 to 225 units, averaging 200 units of the said product. Calculate the following:

(a) Minimum inventory level of Cotton
(b) Maximum inventory level of Nylon
(c) Reorder level of Cotton blend
(d) Average inventory level of Cotton

Answer

In the above problem, the minimum, maximum and average use of material is not given. These have to be calculated from the given data.

We have the maximum, minimum and average production volumes given in units. We also have the per unit consumption of material. The maximum use can be calculated as:

Maximum use = Maximum production in units x units of material required per product

Maximum usage for material Cotton = 225 x 10 = 2,250 kilograms
Maximum usage for material Nylon = 225 x 4 = 900 kilograms
Maximum usage for material Cotton blend = 225 x 6 = 1,350 kilograms

Minimum usage = Minimum production in units x units of material required per product

Minimum usage for material Cotton = 175 x 10 = 1,750 kilograms
Minimum usage for material Nylon = 175 x 4 = 700 kilograms
Minimum usage for material Cotton blend = 175 x 6 = 1,050 kilograms
Average usage = Average production in units x units of material required per product
Average usage for material Cotton = 200 x 10 = 2,000 kilograms
Average usage for material Nylon = 200 x 4 = 800 kilograms
Average usage for material Cotton blend = 200 x 6 = 1,200 kilograms
Calculation of the various inventory levels using the above calculated data:

(a) Minimum inventory level of Cotton = Reorder level – (Average usage x Average lead time)
= 8,000 – (2,000 x 2)
= 4,000 kilograms

(b) Maximum inventory level of Nylon = Reorder level + Reorder quantity – (Minimum usage x Minimum lead time)
= 4,750 +5,000 - (700 x 3)
= 7,650 kilograms

(c) Reorder level of Cotton blend = Maximum usage x Maximum lead time
= 1,350 x 4
= 5,400 kilograms

OR

Reorder level of Cotton blend = Minimum level + (Average usage x Average lead time)
= 2,000 + (1,200 x 3)
= 2,000 + 3,600 = 5,600 kilograms

Continued on the next page

(d) Average inventory level of Cotton = Minimum level +1/2 Reorder quantity
= 4,000 + 1/2 x10, 000
= 4,000 + 5,000
= 9,000 kilograms

OR

Average Inventory level of Cotton $= \dfrac{\text{Maximum level + Minimum level (WN 1)}}{2}$

$= \dfrac{(16{,}250 + 4{,}000)}{2}$ =10,125 kilograms

W1

Maximum level of Cotton = Reorder level + Reorder quantity - (Minimum usage x Minimum lead time)
= 8,000 + 10,000 - (1,750 x 1)
=16,250 kilograms

Please note: average inventory level of cotton comes to two different figures by applying two different formulae.

Test Yourself 6

In Excellent Plc, two components of raw material i.e. plastic and xygon, a chemical, are used for production. The details of the use are given below:

Normal usage 50 units per week of each
Maximum usage 75 units per week of each
Minimum usage 25 units per week of each
Reorder quantity plastic: 300 units; xygon: 500 units
Reorder period plastic: 4 to 6 weeks xygon: 2 to 4 weeks

Required:

Calculate the following for each component:
(a) Reordering level
(b) Minimum level
(c) Maximum level
(d) Average inventory level

Test Yourself 7

Quality Ltd manufactures cookies. Each unit of cookies requires cheese, butter and curd as raw material in the quantities 2 kg, 3 kg and 4 kg respectively. The re-order level of cheese and butter are 6,000 kg and 4,000 kg respectively while minimum level of curd is 1,000 kg. Weekly production of cookies varies from 120 to 200 units, while the weekly average production is 160 units.

The following additional data is available:

	Cheese	Butter	Curd
Reorder quantity (Ltr)	8,000	6,000	8,000
Delivery (in weeks)			
Minimum	1	2	3
Average	2	3	4
[illegible]	3	4	5

Required:

(a) Minimum inventory level of cheese
(b) Maximum inventory level of butter
(c) Reorder level of curd

Answers to Test Yourself

Answer to TY 1

The correct option is D.

Interest cost paid for a loan borrowed for the purchase of inventory is a part of the costs that arise due to the holding of inventory. The salaries of the manager, the watchman guarding the warehouse and the factory supervisor are all fixed costs that will be incurred irrespective of whether or not one holds inventory. Hence these costs are not incurred as a result of holding inventory. Hence, they do not form a part of the inventory holding costs.

Answer to TY 2

The correct option is C.

The calculation of the EOQ will require us to calculate the carrying cost per unit per annum.

The carrying cost is given as Tshs100 cents per bearing per month. Hence the carrying cost per bearing per annum will be calculated as = Tsh100 x 12 months = Tshs1,200

Optimum production run size (EOQ) = $\sqrt{\frac{2DCo}{C_h}}$

$$\sqrt{\frac{2 \times 24{,}000 \text{ bearings} \times \text{Tshs}324{,}000}{\text{Tshs}1{,}200}}$$

= $\sqrt{12{,}960{,}000}$

= 3,600 bearings

Answer to TY 3

In the given problem we have been provided with the discounts that Star Plc. can avail for different ordering quantities.

To determine the optimal quantity to order at which the ordering and holding costs are the minimum, we need to calculate the ordering cost and holding cost of the inventory for different ordering quantities.

The ordering and holding cost for all the ordering quantities is calculated in the table given below:

Ordering quantity (tonnes)	Price per litre (Tshs'000)	No. of orders to be placed	Ordering cost (Tshs'000)	Inventory holding cost (Tshs '000)	Total of ordering and holding costs (Tshs'000)
		(1000/ordering quantity)	(No. of orders to be placed x 150)	(Average inventory) x purchase price per litre x 20%	
150	25	7	1,000	375	1,375
250	24	4	600	600	1,200
500	23	2	300	1,150	1,450
750	20	1	200	1,500	1,700
1,000	18	1	150	1,800	1,950

From the above table, we can infer that the most economical order size is 1,000 litres. So, lots of 1,000 litres should be ordered.

Note: The ordering quantity is divided by 2 to obtain the average inventory in order to calculate the annual holding cost as one does not hold units equal to the optimal re-order quantity at all times.

Answer to TY 4

The given information is

$$EBQ = \sqrt{\frac{2C_oD}{C_h(1 - D/R)}}$$

D = 3,000 units
R = 300 x 50 = 15,000 units
Set up cost: C_o = Tshs500,000
Holding cost per unit per annum = C_h = Tshs2,000

$$\sqrt{\frac{2 \times 3{,}000 \text{ units} \times \text{Tshs}500{,}000}{\text{Tshs}2{,}000 \left(1 - \frac{3{,}000}{15{,}000}\right)}}$$

= 1,369 units.

Answer to TY 5

The correct option is C.

For the inventory model with planned shortages, annual orderings cost = annual holding costs + annual backordering cost

Answer to TY 6

(a) Reorder level = Maximum usage x Maximum lead time
Reorder level for component plastic = 75 units x 6 = 450 units
Reorder level for component xygon = 75 units x 4 = 300 units

(b) Minimum level = Reorder level – (Average usage x Average lead time)
Minimum level for component plastic = 450 units – (50 units x 5) = 200 units
Minimum level for component xygon = 300 units – (50 units x 3) = 150 units

Note: Average usage here refers to the normal usage given.

Working note

$$\text{Average lead time} = \frac{\text{Maximum period} + \text{Minimum period}}{2}$$

$$\text{Average lead time for component plastic} = \frac{(6 + 4)}{2} = 5 \text{ weeks}$$

$$\text{Average lead time for component xygon} = \frac{(2 + 4)}{2} = 3 \text{ weeks}$$

(c) Maximum level = Reorder level + Reorder quantity - (Minimum usage x Minimum lead time)
Maximum level for component plastic = 450 units+ 300 units – (25 units x 4) = 650 units
Maximum level for component xygon = 300 units + 500 units – (25 units X 2) = 750 units

$$\text{(d) Average inventory level} = \frac{\text{Maximum level} + \text{Minimum level}}{2}$$

$$\text{Average inventory level for component plastic} = \frac{650 \text{ units} + 200 \text{ units}}{2} = 425 \text{ units}$$

$$\text{Average inventory level for component xygon} = \frac{750 \text{ units} + 150 \text{ units}}{2} = 450 \text{ units}$$

Answer to TY 7

(a) Minimum inventory level of cheese

= Reorder Level – (Normal or average usage x Average delivery time)
= 6,000 – (160 units x 2 kg x 2 weeks)
= 6,000 – 640 kg
= 5,360 kg

(b) Maximum inventory level of butter

= Reorder level + Reorder quantity – Minimum consumption to obtain delivery
= 4,000 kg + 6,000 kg – (120 unit x 3 kg x 2 weeks)
= 10,000 kg – 720 kg
= 9,280 kg

(c) Reorder level of curd

= Maximum reorder period x Maximum usage
= 5 weeks x 200 units x 4 kg
= 4,000 kg

Self-Examination Questions

Question 1

Minimum inventory level + (Average consumption x Average delivery period) is equal to:

A Danger level
B Maximum level
C Economic quantity
D Reorder level

Question 2

____________ is the level at which an immediate purchase needs to be made to avoid a stock-out situation.

A Minimum level
B Danger level
C Optimum level
D Constant level

Question 3

Symphony Industries wants to purchase oil tins. Their annual requirement is 5,000 tins. Each order costs Tshs250,000. The cost of carrying inventory is Tshs20,000.

Economic Order Quantity (EOQ) will be:

A 500 tins
B 354 tins
C 400 tins
D 325 tins

Question 4

Using the information in Self Examination Question 3, if purchase cost per oil tin is Tshs100,000, what will be the total annual cost?

A Tshs500,000,000
B Tshs503,351,000
C Tshs507,071,073
D Tshs503,540,000

Question 5

The monthly requirement of a certain raw material for Libra Ltd is 400 units. The purchase price per unit is Tshs15,000. Holding cost is 10% of the purchase price, and the ordering cost is Tshs100,000 per order. The total inventory cost for the current year is Tshs75,000,000 including purchase cost, holding cost and ordering cost. If EOQ is used, in comparison to the current ordering policy the amount saved will be:

A Tshs800,000
B Tshs8,000,000
C Tshs180,000
D Tshs1,800,000

Question 6

Neptune Ltd is a fabric manufacturing company.

Its annual demand for synthetic fibre will be	45,000 metres
Order placing cost per order	Tshs25,000
Cost per meter of synthetic fibre	Tshs2,000
Storage cost	10% on average inventory

Required:

(a) By using the above information, calculate EOQ for Neptune Ltd.
(b) Also calculate number of orders to be placed in a year.

Answers to Self Examination Questions

Answer to SEQ 1

The correct option is D.

The formula for the reorder level is:

Reorder level= Minimum level + (Average usage x Average lead time)

Answer to SEQ 2

The correct option is B.

The danger level is the volume of inventory at the minimum level, and is a trigger for an immediate action of purchase to avoid a stock out situation.

Answer to SEQ 3

The correct option is B.

Here, D = annual demand = 5,000 units
C_o = ordering costs = Tshs250,000
C_h = holding costs = Tshs20,000

Optimum purchase order size (EOQ) = $\sqrt{\frac{2DC_o}{C_h}}$

$$\sqrt{\frac{2 \times \text{Tshs}250{,}000 \times 5{,}000}{\text{Tshs}20{,}000}}$$

= $\sqrt{125{,}000}$

= 353.55 tins

= 354 tins

Answer to SEQ 4

The correct option is C.

Total annual cost = $P + \left[C_o \times \frac{D}{Q}\right] + \left[C_h \times \frac{Q}{2}\right]$

$$(\text{Tshs}100{,}000 \times 5{,}000) + \left[\text{Tshs}250{,}000 \times \frac{5{,}000}{354}\text{ tins}\right] + \left[\text{Tshs}20{,}000 \times \frac{354}{2}\text{ tins}\right]$$

= Tshs500,000,000 + Tshs3,531,073 + Tshs3,540,000
= Tshs507,071,073

Answer to SEQ 5

The correct option is D.

Saving is always a positive amount.

$$EOQ = \sqrt{\frac{2C_oD}{C_h}}$$

$$\sqrt{\frac{2 \times Tshs100{,}000 \times 4{,}800}{Tshs15{,}000 \times 10\%}}$$

$$= \sqrt{\frac{2 \times 100 \times 4{,}800}{15 \times 10\%}}$$

= 800 units

The savings are calculated as follows:

	Tshs
Purchase costs (4,800 units x Tshs15,000)	72,000,000
Order costs (4,800/800 x 100,000)	600,000
Holding costs (800/2 x 15,000 x 10%)	600,000
Total costs	73,200,000
Less: Original inventory costs	75,000,000
Savings	1,800,000

Note: Monthly demand is 400 units. Therefore annual demand is calculated as 4,800 units.

Answer to SEQ 6

(a) The calculation of EOQ is as follows:

$$EOQ = \sqrt{\frac{2C_oD}{C_h}}$$

Where,

C_o = Cost of ordering per order / consignment from supplier

C_h = Cost of holding per unit of inventory per annum / time period

D = Total demand during the period

$$EOQ = \sqrt{\frac{2 \times 45{,}000 \times 25{,}000}{(Tshs2{,}000 \times 10\%)}}$$

$$EOQ = \sqrt{\frac{2{,}250{,}000}{0.20}}$$

$$EOQ = \sqrt{11{,}250{,}000}$$

= 3354

No. of orders to be placed in a year = Consumption of materials per annum / EOQ

$$= \frac{45{,}000}{3354}$$

= 13 orders per year

(b) 13 orders need to be placed per year with an order size of 3354 meters to keep the ordering and the holding cost at the minimum level.

QUEUING MODELS

12

Study Guide 12 -A: COMPONENTS OF A QUEUE SYSTEM, ARRIVALS, and SERVICES IN M/M/1 QUEUE

Get Through Intro

Queuing models provide the analyst with a powerful tool for designing and evaluating the performance of queuing systems. Queuing Model is a model used for solving a problem where certain service facilities have to provide service to its customers, so as to avoid lengthy waiting line or queue, so that customers will get satisfaction from effective service and idle time of service facilities are minimized is waiting line model or queuing model.

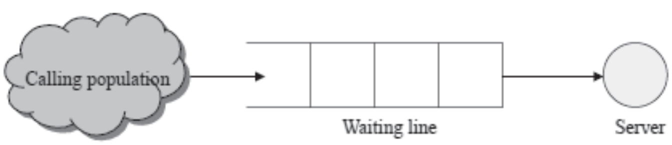
Whenever customers arrive at a service facility, some of them have to wait before they receive the desired service. It means that the customer has to wait for his/her turn, may be in a line. Customers arrive at a service facility with several queues, each with one server (sales checkout counter). The application of queuing models is highly demanding in almost all fields of life. Delays and queuing problems are most common features not only in our daily-life situations such as at a bank or postal office, at a ticketing office, in public transportation or in a traffic jam but also in more technical environments, such as in manufacturing, computer networking and telecommunications.
Sometimes, insufficiencies in services also occur due to an undue wait in service may be because of new employee. Delays in service jobs beyond their due time may result in losing future business opportunities. Queuing theory is the study of waiting in all these various situations. It uses queuing models to represent the various types of queuing systems that arise in practice. The models enable finding an appropriate balance between the cost of service and the amount of waiting.

Learning Outcomes

a) Identify the components of a queue system.
b) Determine the characteristics values of a queue system.
c) Determine the cost of a queue.

Simulation is often used in the analysis of queueing models.

Figure xxx: A simple but typical queueing model



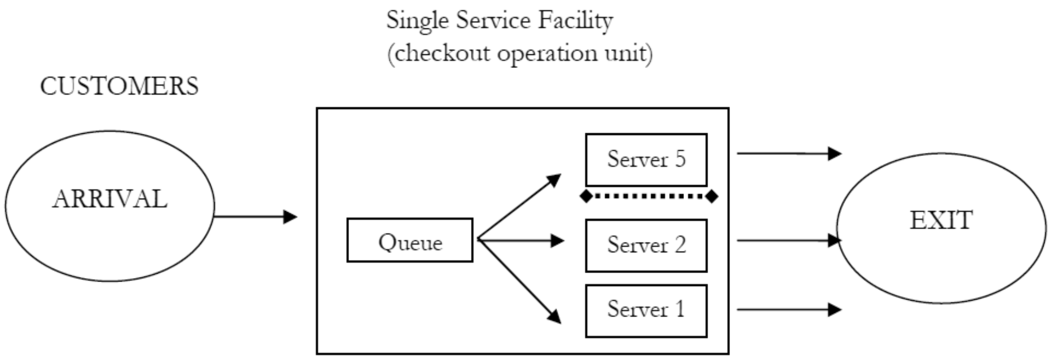
Queueing models provide the analyst with a powerful tool for designing and evaluating the performance of queueing systems. Typical measures of system performance such as: Server utilization, length of waiting lines, and delays of customers. For relatively simple systems: compute mathematically and For realistic models of complex systems: simulation is usually required

Basic Queuing Process

Customers requiring service are generated over time by an input source. The required service is then performed for the customers by the service mechanism, after which the customer leaves the queuing system. We can have following two types of models (Nafees, 2007):

a) One model will be as Single-queue Multiple-Servers model (Figure 1)

Figure 1: Single Stage Queuing Model with Single-Queue and Multiple Parallel Servers



b) The second one is Multiple-Queues, Multiple-Servers model (Figure 2)

Figure 2: Single Stage Queuing Model with Multiple Queues and Multiple Parallel Servers

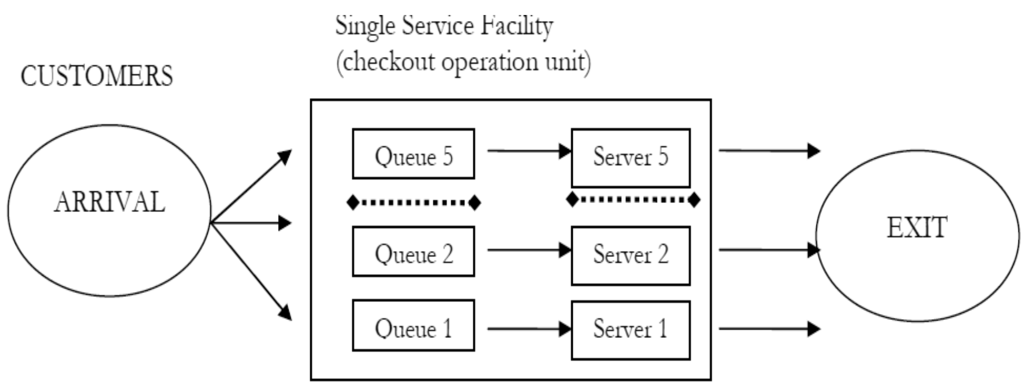
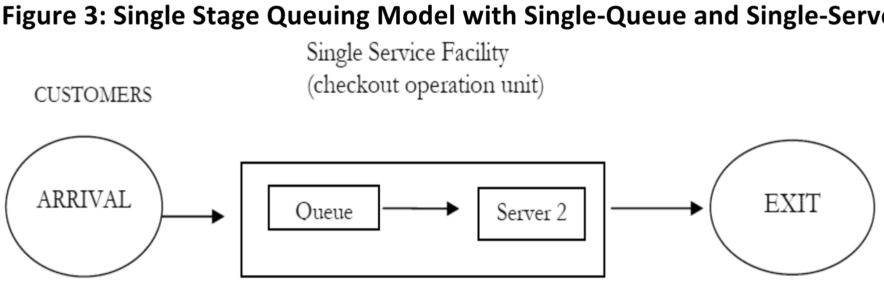


Figure 3: Single Stage Queuing Model with Single-Queue and Single-Server



In these models, three various sub-processes may be distinguished:

i) **Arrival Process:** includes number of customers arriving, several types of customers, and one type of customers' demand, deterministic or stochastic arrival distance, and arrival intensity. The process goes from event to event, i.e. the event "customer arrives" puts the customer in a queue, and at the same time schedules the event "next customer arrives" at some time in the future.

ii) **Waiting Process:** includes length of queues, servers' discipline (First In First Out). This includes the event "start serving next customer from queue" which takes this customer from the queue into the server, and at the same time schedules the event "customer served" at some time in the future.

iii) **Server Process:** includes a type of a server, serving rate and serving time. This includes the event "customer served" which prompts the next event "start serving next customer from queue".

Characteristics of Queueing Systems

Any queuing system has three parts:

(1) the arrivals or inputs to the system (sometimes referred to as the calling population)
(2) the queue or the waiting line itself, and
(3) the service facility.

Key elements of queueing systems

- **Customer:** refers to anything that arrives at a facility and requires service, e.g., people, machines, trucks, emails, packets, frames.
- **Server:** refers to any resource that provides the requested service, e.g., repairpersons, machines, runways at airport, host, switch, router, disk drive, algorithm.

These three components have certain characteristics that must be examined before mathematical queuing models can be developed.

i) **Arrival Characteristics**

The input source that generates arrivals or customers for the service system has three major characteristics. It is important to consider the size of the calling population, the pattern of arrivals at the queuing system, and the behaviour of the arrivals.

Table xxx: Elements of a Queuing System

System	**Customers**	**Server**
Reception desk	People	Receptionist
Hospital	Patient	Nurse
Airport	Airplanes	Runways
Production line	Cases	Case-parker
Road Network	Cars	Traffic lights
Grocery	Shoppers	Checkout station
Computer	Jobs	CPU, Disks, CD
Network	Packets	Router

ii) **Calling Population**

The Size of the Calling Population are considered to be either unlimited (essentially infinite) or limited (finite). When the number of customers or arrivals on hand at any given moment is just a small portion of potential arrivals, the calling population is considered unlimited. For practical purposes, examples of unlimited populations include cars arriving at a highway tollbooth, shoppers arriving at a supermarket, or students arriving to register for classes at a large university.

Most queuing models assume such an infinite calling population. When this is not the case, modeling becomes much more complex. An example of a finite population is a shop with only eight machines that might break down and require service.

Calling population: the population of potential customers may be assumed finite or infinite.

- **Finite population model:** if arrival rate depends on the number of customers being served and waiting, e.g., model of one corporate jet, if it is being repaired, the repair arrival rate becomes zero.

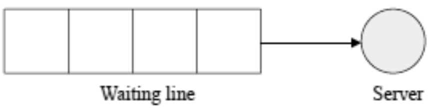
- **Infinite population model:** if arrival rate is not affected by the number of customers being served and waiting, e.g., systems with large population of potential customers.

iii) **System Capacity**

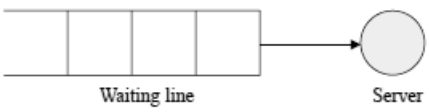
System Capacity: a limit on the number of customers that may be in the waiting line or system.

- Limited capacity, e.g., an automatic car wash only has room for 10 cars to wait in line to enter the mechanism.

- If system is full no customers are accepted anymore



- Unlimited capacity, e.g., concert ticket sales with no limit on the number of people allowed waiting to purchase tickets.



iv) **Arrival Process**

- For infinite-population models: In terms of inter-arrival times of successive customers.
- Arrival types:
  - Random arrivals: interarrival times usually characterized by a probability distribution.
  - Most important model: Poisson arrival process (with rate λ), where a time represents the interarrival time between customer n-1 and customer n, and is exponentially distributed (with mean 1/λ).
  - Scheduled arrivals: interarrival times can be constant or constant plus or minus a small random amount to represent early or late arrivals.
    - Example: patients to a physician or scheduled airline flight arrivals to an airport
- At least one customer is assumed to always be present, so the server is never idle, e.g., sufficient raw material for a machine.
- For finite-population models:
  - Customer is pending when the customer is outside the queueing system, e.g., machine-repair problem: a machine is "pending" when it is operating, it becomes "not pending" the instant it demands service from the repairman.
- Runtime of a customer is the length of time from departure from the queueing system until that customer's next arrival to the queue, e.g., machine-repair problem, machines are customers, and a runtime is time to failure (TTF).

Waiting Line Characteristics

The waiting line itself is the second component of a queuing system. The length of a line can be either *limited* or *unlimited*. A queue is limited when it cannot, by law of physical restrictions, increase
to an infinite length. This may be the case in a small restaurant that has only 10 tables and can serve no more

than 50 diners an evening. Analytic queuing models are treated in this chapter under an assumption of **unlimited queue length**. A queue is unlimited when its size is unrestricted, as in the case of the tollbooth serving arriving automobiles. A second waiting line characteristic deals with **queue discipline**. This refers to the rule by which customers in the line are to receive service

Most systems use a queue discipline known as the **first-in, first-out** (FIFO) rule.

- In a hospital emergency room or an express checkout line at a supermarket, however, various assigned priorities may pre-empt FIFO.
- Patients who are critically injured will move ahead in treatment priority over patients with broken fingers or noses.
- Shoppers with fewer than 10 items may be allowed to enter the express checkout queue but are *then* treated as first come, first served.
- Computer programming runs are another example of queuing systems that operate under priority scheduling. In most large companies, when computer-produced pay checks are due out on a specific date, the payroll program has highest priority over other runs.

Queue Behaviour and Queue Discipline

Queue behaviour: the actions of customers while in a queue waiting for service to begin, for example:

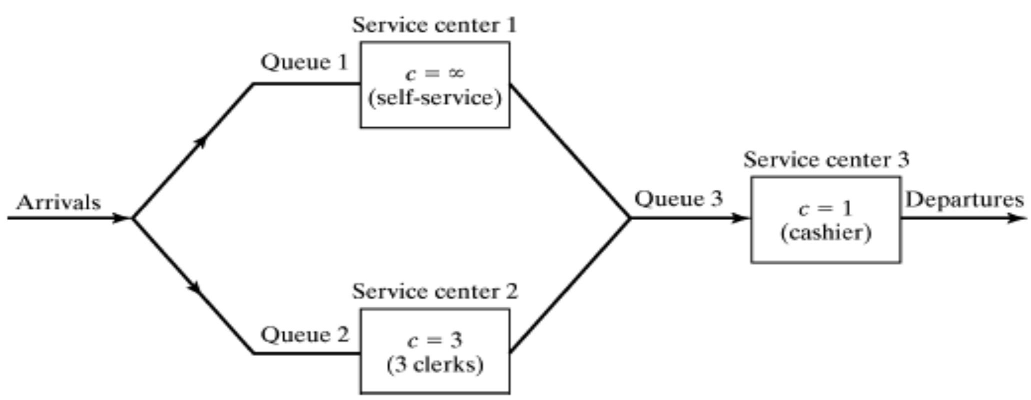
- Balk: leave when they see that the line is too long
- Renege: leave after being in the line when its moving too slowly
- Jockey: move from one line to a shorter line

Queue discipline: the logical ordering of customers in a queue that determines which customer is chosen for service when a server becomes free, for example:

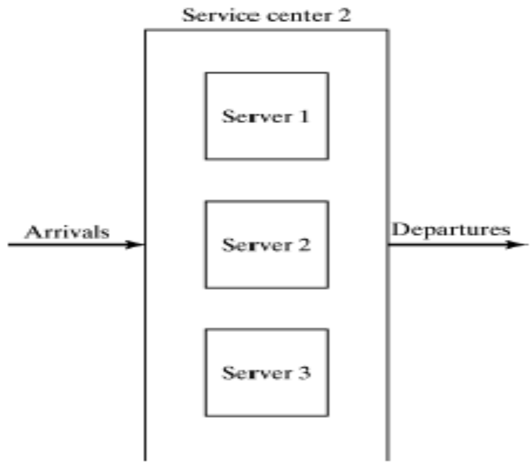
- First-in-first-out (FIFO)
- Last-in-first-out (LIFO)
- Service in random order (SIRO)
- Shortest processing time first (SPT)
- Service according to priority (PR)

Queuing System: Example 2

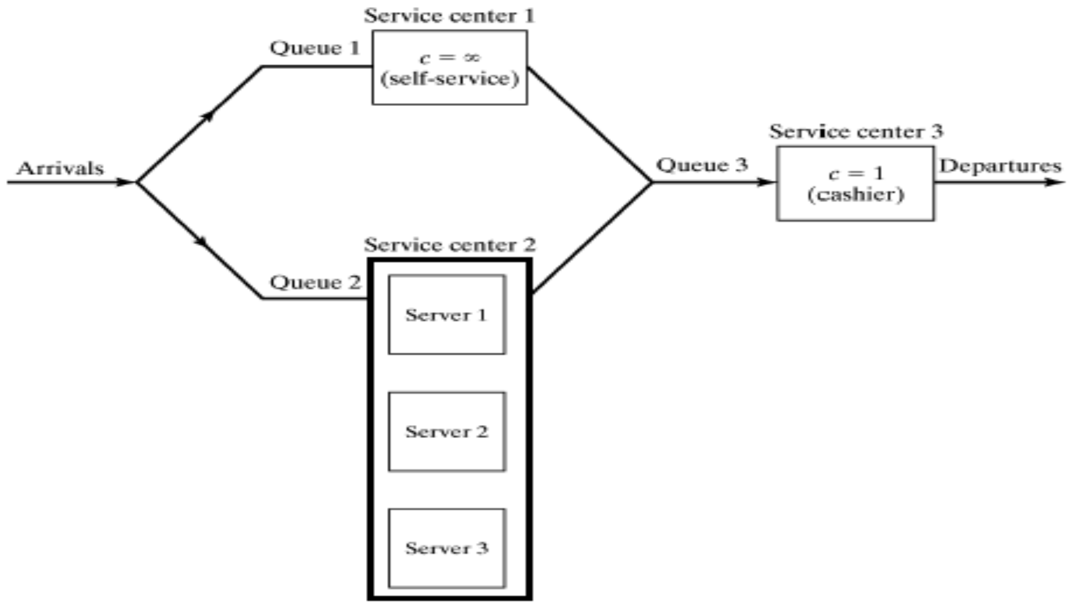
Example: consider a discount warehouse where customers may serve themselves before paying at the cashier (service center 1) or served by a clerk (service center 2).



Wait for one of the three clerks:

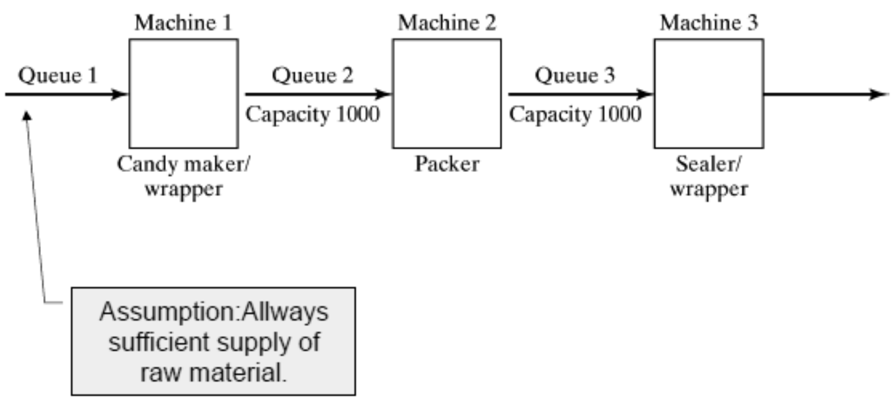


Batch service (a server serving several customers simultaneously), or customer requires several servers simultaneously.



Sweets (Cundy) production line

- Three machines separated by buffers
- Buffers have capacity of 1000 candies



Service Facility Characteristics

The third part of any queuing system is the service facility. It is important to examine two basic properties:

(1) the configuration of the service system and

(2) the pattern of service times.

- **Basic Queuing System Configurations**

Service systems are usually classified in terms of their number of channels, or number of servers, and number of phases, or number of service stops, that must be made.

- A **single-channel system**, with one server, is typified by the drive-in bank that has only one open teller, or by the type of drive-through fast-food restaurant.
- If, on the other hand, the bank had several tellers on duty and each customer waited in one common line for the first available teller, we would have a **multichannel system** at work.

Many banks today are multichannel service systems, as are most large barber shops, supermarkets counters and many airline ticket counters.

Pattern of Arrivals at The System

Customers either arrive at a service facility according to some known schedule (for example, one patient every 15 minutes, or one student for advising every half hour) or else they arrive randomly. Arrivals are considered random when they are independent of one another and their occurrence cannot be predicted exactly. Frequently in queuing problems, the number of arrivals per unit of time can be estimated by a probability distribution known as the **Poisson distribution**.

Thus, a single channel model with Poisson arrivals and exponential service times would be represented by: M/M/1

When a second channel is added, we would have M/M/2. An M/M/2 model has Poisson arrivals, exponential service times, and two channels.

If there are m distinct service channels in the queuing system with Poisson arrivals and exponential service times, the Kendall notation would be M/M/m. A three-channel system with Poisson arrivals and constant service time would be identified as M/D/3. A four-channel system with Poisson arrivals and service times that are normally distributed would be identified as M/G/4.

The basic three-symbol **Kendall notation** is in the form: **Arrival distribution/ Service time distribution/Number of service channels open.** Where specific letters are used to represent probability distributions. The following letters are commonly used in Kendall notation:

- M = Poisson distribution for number of occurrences (or exponential times)
- D = constant (deterministic) rate
- G = general distribution with mean and variance known.

Queuing Equations

We let,

λ= mean number of arrivals per time period (for example, per hour)
: = mean number of people or items served per time period
When determining the arrival rate **(λ)** and the service rate (**:**) the same time must be used. For example, if **λ** is the average number of arrivals per hour, then **:** must indicate the average number that could be served per hour.

These seven queuing equations for the single-channel, single-phase model describe the important operating characteristics of the service system.
The queuing equations follow.

i) The average number of customers or units in the system, L, that is, the number in line plus the number being served: $L = \dfrac{\lambda}{\lambda - \mu}$

ii) The average time a customer spends in the system, W, that is, the time spent in line plus the time spent being served: $w = \dfrac{1}{\mu - \lambda}$

iii) The average number of customers in the queue, $L_q = \dfrac{\lambda^2}{\mu(\mu - \lambda)}$

iv) The average time a customer spends waiting in the queue, $W_q = \dfrac{\lambda}{\mu(\mu - \lambda)}$

v) The utilization factor for the system, Δ (the Greek lowercase letter rho), that is, the probability that the service facility is being used: $\rho = \dfrac{\lambda}{\mu}$

vi) The percent idle time, P_o that is, the probability that no one is in the system: $P_0 = 1 - \dfrac{\lambda}{\mu}$

vii) The probability that the number of customers in the system is greater than K, $P_{n>K}$

$$P_{n>k} = \left(\frac{\lambda}{\mu}\right)^{k+1}$$

The simplest queuing model is the M/M/1, for which key formulas are:

$P_0 = 1 - \rho$	$L_s = \dfrac{\lambda}{\mu - \lambda}$	$W_s = \dfrac{1}{\mu - \lambda}$
$P_n = P_0\rho^n$	$L_q = \dfrac{\rho\lambda}{\mu - \lambda} = \dfrac{\rho^2}{1 - \rho}$	$W_q = \dfrac{\rho}{\mu - \lambda}$

Example 1

Customers arrive at a bank at a rate of 30 per hour. Arrivals are random and service time is exponential, so that the *M*/*M*/1 model applies. The clerk's service time is 90 seconds. Therefore $\lambda = 30$ and $\mu = 45$, and as $\rho = \lambda/\mu = 30/45 = ⅔ < 1$, the system has a steady state.

The probability of an idle bank teller is $P_0 = 1 - \rho = ⅓$. The probability that at least two customers are waiting is $P_3 + P_4 + P_5 + \ldots = 1 - P_0 - P_1 - P_2 = 1 - ⅓ - ⅓(⅔)^1 - ⅓(⅔)^2 \cong = \dfrac{8}{27} \approx 0.2963$ or slightly less than one third.

On average, there are $L_q = 1.3333$ customers waiting in line and the average time for a customer in the system is $W_s = 1/15$ hour $= 4$ minutes.

Suppose that in an $M/M/1$ system with $\lambda = 20$, the decision maker specifies that the probability of three or more customers in the system should not exceed 95%. This probability is $1 - P_0 - P_1 - P_2 = 1 - (1-\rho) - \rho(1-\rho) - \rho^2(1-\rho) = 1 - 1 + \rho - \rho + \rho^2 - \rho^2 + \rho^3 = \rho^3$, and since it should not exceed 95%, we obtain $p^3 = \frac{\lambda^3}{u^3} \leq 0.95$ or $\mu^3 > 8{,}000/.95$ or $\mu > 20.3449$.

Example 2

The Directorate of Information and Communication Technology at Mzumbe University has a help desk to assist students working on computer spreadsheet assignments. The students patiently form a single line in front of the desk to wait for help. Students are served based on a first-come, first-served priority rule. On average, 15 students per hour arrive at the help desk. Student arrivals are best described using a Poisson distribution. The help desk server can help an average of 20 students per hour, with the service rate being described by an exponential distribution. Calculate the following operating characteristics of the service system.

a) The average utilization of the help desk server
b) The average number of students in the system
c) The average number of students waiting in line
d) The average time a student spends in the system
e) The average time a student spends waiting in line
f) The probability of having more than 4 students in the system

Hints for solving this kind of questions

The key to solving queuing problems is to identify the mean arrival rate of customers and the mean service rate. In this case, on average, 15 customers arrive each hour. On average, the consultant can serve 20 customers per hour. Once you have established these values, you merely plug them into the appropriate formula**.**

Solution

a) Average utilization: $P = \frac{\lambda}{\mu} = \frac{15}{20}$ =0.75 0r 75%.
b) Average number of students in the system: $L = \frac{\lambda}{\mu - \lambda} = \frac{15}{20-15} = 3$ Students
c) Average number of students waiting in line: $L_Q =$ pL = 0.75 x 3 = 2.25 students
d) Average time a student spent in the system: = 0.2 hours or 12 minutes.
e) Average time a student spent waiting in line: $W_Q =$ pW = 0.75 x (0.2) = 0.15 hours, or 9 minutes
f) The probability that there are more than four students in the system equals one minus the probability that there are four or fewer students in the system. We use the following formula.

Problem-Solving Tip: Any term raised to the zero power is equal to 1.

$$P = 1 - \sum_{n=0}^{4} p_n = 1 - \sum_{n=0}^{4} (1-p)pn$$

$= 1- 0.25\ (1+ 0.75 + 0.75^2+ 0.75^3+ 0.75^4)$

$= 1- 0.7626 = 0.2374$

or a 0.2374 (23.74 percent) chance of having more than four students in the system.

Example 3

A new shopping mall is considering setting up an information desk manned by one employee. Based upon information obtained from similar information desks, it is believed that people will arrive at the desk at a rate of 20 per hour. It takes an average of 2 minutes to answer a question. It is assumed that the arrivals follow a Poisson distribution and answer times are exponentially distributed.

(a) Find the probability that the employee is idle.

(b) Find the proportion of the time that the employee is busy.

(c) Find the average number of people receiving and waiting to receive some information.

(d) Find the average number of people waiting in line to get some information.

(e) Find the average time a person seeking information spends in the system.

(f) Find the expected time a person spends just waiting in line to have a question answered (time in the queue).

Solution:

(a) $P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{20}{30} = 0.33 => 33\%$

(b) $p = \frac{\lambda}{\mu} = 0.66$

(c) $L_s = \frac{\lambda}{\lambda - \mu} = \frac{20}{30-20} = 2$ people

(d) $L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{20^2}{30(30-20)} = 1.33$ people

(e) $W_s = \frac{1}{\lambda - \mu} = \frac{1}{30-20} = 0.10$ hours

(f) $W_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{20}{30(30-20)} = 0.0667$ hours

Example 4

Assume that the information desk employee in Example 3 Above earns TZS 10,000.00 per hour. The cost of waiting time, in terms of customer unhappiness with the mall, is TZS 12,000.00 per hour of time spent waiting in line. Find the total expected costs over an 8-hour day.

The average person waits 0.0667 hours and there are 160 (20 arrivals X 8 hours) arrivals per day.

Therefore: Total waiting time = 160 x 0.0667 = 10.67 hours

Total cost for waiting = Total waiting time X Cost per hour = 10.67 X TZS. 12,000.00 = TZS.128, 040.00 per day.

Salary cost = 8 hours X TZS.10,000.00 = TZS. 80,000.00

Total cost = Salary cost + Waiting cost = TZS.80,000.00 + TZS.128, 040.00 = TZS. 208,040.00 per day.

Test yourself 1

1. What is the Poisson distribution?

2. Three students arrive per minute at a coffee machine that dispenses exactly four cups per minute at a constant rate. Describe the system parameters.

Solution to test Yourself Questions

1. Simply a discrete probability distribution that happens to generate numbers that strongly resembles the numbers we see when we collect arrival data over a long period of time. Since it is a lot easier to use a probability distribution to look at arrivals than it is to collect and examine a couple of months' worth of arrival data, we tend to use the distribution
2. $L_q = 1.125$ people in the queue on average

   Wq = 0.375 minutes in the queue waiting

   Ls = 1.87 people in the system

   Ws = 0.625 minutes in the system

QUEUING MODELS

12

Study Guide 12 -B: M/M/2 MODEL

Get Through Intro

The Queuing model is commonly labelled as M/M/c/K, where first M represents Markovian exponential distribution of inter-arrival times, second M represents Markovian exponential distribution of service times, c (a positive integer) represents the number of servers, and K is the specified number of customers in a queuing system. This general model contains only limited number of K customers in the system. However, if there are unlimited number of customers exist, which means K = Q, then our model will be labeled as M/M/c.

Learning Outcomes

a) Identify the components of a queue system with more than one service point.
b) Determine the characteristics and cost of a queue system with more than one service point.
c) Apply the concept of queue system in accounting and business.

Parameters in Queuing Models (Multiple Servers, Multiple Queues Model) are:

i) n = Number of total customers in the system (in queue plus in service)
ii) c = Number of parallel servers (Checkout sales operation units)
iii) λ= Arrival rate (1 / (average number of customers arriving in each queue in a system in one hour))
iv) := Serving rate (1 / (average number of customers being served at a server per hour))
v) c:= Serving rate when c > 1 in a system
vi) Δ= System intensity or load, utilization factor (=λ /(cμ)) (the expected factor of time the server is busy that is, service capability being utilized on the average arriving customers)

Departure and arrival rate are state dependent and are in steady-state (equilibrium between events) condition.

Notations & their Description for single queue and parallel multiple servers model (Figure 1) assuming the system is in steady-state condition are:

i) P0 = Steady-state Probability of all idle servers in the system i.e

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{\gamma^n}{n!} + \frac{\gamma^c}{c!\,(1-\rho)} \right]^{-1} \quad \text{Where} \quad \gamma = \frac{\lambda}{\mu}$$

ii) Pn= Steady-state Probability of exactly n customers in the system

$$P_n = \frac{\lambda^n}{c!c^{n-c}\mu^n} P_0 \quad n > c$$

iii) Lq = Average number of customers in the waiting line (queue) = $= \dfrac{\gamma^c \rho}{(c)!\,(1-\rho)^2} \times P_0$

iv) Wq = Average waiting time a customer spends waiting in line excluding the service time

$$= \frac{L_q}{\lambda}$$

There are no predefined formulas for networks of queues, i.e. multiple queues such as in Figure 2. A complexity of the model is the main reason for that. Therefore, we use notations and formulas for single queue with parallel servers. In order to calculate estimates for multiple queues multiple servers' model, one could run simulation.

Queuing Simulation

The queuing system is when classified as M/M/c with multiple queues where number of customers in the system and in a queue is infinite, the solution for such models are difficult to compute. When analytical computation of T is very difficult or almost impossible, a Monte Carlo simulation is appealed in order to get estimations. A standard Monte Carlo simulation algorithm fix a regenerative state and generate a sample of regenerative cycles, and then use this sample to construct a likelihood estimator of state.

Discrete models deal with system whose behavior changes only at given instants. A typical example occurs in waiting lines where we are interested in estimating such measures as the average waiting time or the length of the waiting line. Such measures occur only when the customer enters or leaves the system.

The instants at which changes in the system occurs identify the model's events, e.g. arrival and departure of the customers. The arrival events are separated by the 'inter-arrival time' (the interval between successive arrivals), and the departure events are specified by the service time in the facility. The fact that these events occur at discrete points is known as "**Discrete-event Simulation**."

When the interval between successive arrivals is random then randomness arises in simulations. The time t between customers' arrivals is represented by an exponential distribution; to generate the arrival times of the next customers from this distribution, we have t= - (1/:) ln(1-R) Where R = random number. (1-R) is a compliment of R, where R can be replaced by (1- R) with R.

Example

The parameters and corresponding characteristics in Queuing Model M/M/2, assuming system is in steady-state condition (Queuing model 1 in Figure 1), are:

i) c = number of servers = 2
ii) λ = arrival rate = 98 customers per hour
iii) : = serving rate = 55 customers per server per hour
iv) c:= (2) (55) = 110 (service rate for 2 servers)
v) Δ = λ /(c/:) = 98 / 110 = 0.8909
vi) (= λ /: = 1.7818

Overall system utilization =Δ = 89.09 %

The probability that all servers are idle (Po) = 0.5769

Average number of customers in the queue (Lq) = $\dfrac{\gamma^c \rho}{(c)!\,(1-\rho)^2} \times P_0 = 6.8560$

Average time customer spends in the queue (Wq) = Lq/ λ = 0.0700 hours

.

Interpretation of results for queuing model 1

The performance of the sales checkout service on weekday is sufficiently good. We can see that the probability for servers to be busy is 0.8909, i.e. 89.09%. The average number of customers waiting in a queue is Lq = 6.8560 customers per 2-server. The waiting time in a queue per server is Wq = 4.2 min which is normal time in a busy server. This estimate is not realistic as the model shows that the customers make a single queue and choose an available server. Hence, we can consider each server with a queuing model as a single-server single-queue model to get the correct estimate of the length of queue. M/M/1 queue is a useful approximate model when service times have standard deviation approximately equal to their means.

SIMULATION

13

Get Through Intro

Risk analysis is a part of every decision we make. For a multinational company, which has a business set-up in many countries and also a wide range of products, uncertainty and ambiguity in key factors that affect key decisions lead to business challenges. Therefore, in order to run the business successfully, the management would like to accurately estimate the probabilities of uncertain events; for example, what is the probability that the introduction of a new product in the market will generate profits? Will the company be able to supply the product based on the demand pattern?

Simulation models help to provide the decision maker with a range of possible outcomes and the probability with which they occur. Using these simulation models, the decision maker can assess the impact of risk and take an informed decision under uncertainty.

Simulation models are used by various companies in the field of finance, project management, engineering, research and development, insurance, oil and gas, transportation, etc.

In this Study Guide, we will understand the application of simulation techniques in a few of these areas.

Learning Outcomes

a) Understand the meaning of simulation, the simulation process and its advantages and disadvantages.
b) Understand the Monte Carlo simulation model and the method of constructing it.
c) Understand the use of the simulation model for solving queuing problems.
d) Apply simulation techniques to accounting and business situations.

1. Understand the meaning of simulation, the simulation process and its advantages and disadvantages.

[Learning Outcome a]

1.1 Meaning of simulation

Definition

Simulation is a quantitative procedure which imitates the operation of a real world process or system over a period of time.

Simulation can be done either manually or on a computer. Simulation helps in developing a model of that process / system and then conducting a series of organised trial and error experiments to predict the behaviour of the process / system over time.

In other words, simulation involves imitation of some real thing or process. It involves representing certain key characteristics and behaviours of a selected physical and abstract system. This is generally done by modelling the situation in order to represent a real life situation.

In certain cases, it might not be possible to formulate the entire problem or solve it through mathematical models. In such cases, simulation proves to be the most suitable method, which suggests an optimal solution to take decisions.

Simulations can be used to solve a wide variety of business problems, which include:

(a) Inventory control
(b) Queuing
(c) Production planning

Example

A bank started a service for its customers to pay utility bills through their branch. To pay the bills, the customers have to deposit cheques along with a copy of their bills. The bank personnel then issue acknowledgement receipts to the customers. The number of customers visiting the bank to pay their bills will determine the number of customer service desks to be set up by the branch. In such cases, simulation models can be helpful in simulating the average wait time of customers, average transactions processed during the day, etc.

1.2 Process of Simulation

The process of simulation is very simple and easily understandable. All we need to do is construct a model of the actual situation and then estimate the results by replacing the variable factors with sample data.

The following are the steps that are usually followed in a simulation process:

(a) The problem is defined.
(b) The variables of the model that have a logical relationship with one another are introduced and a suitable
(c) model is constructed.
(d) After developing a desired model, each alternative is evaluated by generating a series of values of the
(e) random variable, and the behaviour of the system is observed.
(f) Lastly, the results are examined and the best alternative is selected.

The whole process has been shown below with the help of a flow chart.

Diagram 1: Simulation Process

1. •Define the problem you intend to simulate
2. •Introduction of variables
3. •Construction of simulation model
4. •Testing of the model by replacing variables with values
5. •Running the simulation
6. •Examination of Results
7. •Repeating the simulation if results not acceptable
8. •Selecting the best alternative

1.3 Advantages of a simulation model

The following are the advantages of the simulation model:

1. Simulation can analyse complex and large practical problems when it is not possible to solve them through mathematical methods.

Example

Snowballs Ltd is a very popular brand of ice cream and it has several retail outlets throughout the globe which serve ice creams in different flavours. The management of Snowballs wants to forecast the sales for the Strawberry flavour so that it can procure raw materials accordingly. This is a bit tricky situation as the demand for the strawberry flavour has been erratic.

In such a situation, Snowballs can simulate the sale of strawberry flavoured ice-cream for a particular month with the help of a simulation model. This will help them to plan out their production and utilise resources more efficiently.

---

2. Simulation is flexible; hence changes in the system variables can be made to select the best solution among the various alternatives.

Example

Arnie, the owner of a fast food restaurant, is thinking of starting a separate coffee shop in his restaurant. The manager plans for one service counter exclusively for the coffee shop customers. According to a market study, the approximate arrival times of customers are projected.

Now Arnie wants to know the average waiting time of customers to decide whether one service counter would be enough or two are needed.

In this case, the problem can be solved easily using the simulation technique, as the variables in simulation are flexible and can be replaced with random numbers to depict a real life situation. (Random numbers are explained in detail in Learning Outcome 2)

So, Arnie can develop a simulation model for the activities at the coffee shop and estimate the average waiting time of the customers.

---

3. In simulation, the experiments are carried out in isolation and does not disturb the ongoing operations of the real life systems.

Example

If we want to estimate the average waiting time of a customer at a grocery shop, we would have to fiddle with the actual system to determine that. But with the help of the simulation technique, we can construct a simulation model of the situation and use random numbers to get the desired information.

4. Policy decisions can be made much faster by knowing the options well in advance and by reducing the risks involved in experimenting with the real system.

Example

A baker supplies 200 loaves of bread every day to the outlet situated in the city. However, the production of bread varies depending upon the availability of raw materials and labour, for which the probability distribution of production by observation has been made.

The baker wants to find out the average loaves of bread produced in excess of the requirement and the average shortage of loaves at the outlet. The baker can then estimate this by constructing a model of the problem and replacing the variables with random numbers, instead of experimenting with the actual labour and raw material supplies.

This process would be very fast and less risky.

5. A simulation model can help in early detection of bottlenecks in a real life system. This can beneficial in determining the delay or lag in getting information, material or any other inputs for which the subsequent processes are dependent on.

6. A simulation model is just an experiment and does not require organisations to commit the resources. This can be cost effective if the company does not wants to spend much on research activities.

1.4 Disadvantages of a simulation model

The following are the disadvantages of the simulation model:

1. Simulation models may not provide the exact or the optimal solutions for the processes / systems.

2. It may take a long time to develop a good simulation model.

3. In certain cases, simulation models can be very expensive.

4. The decision-maker must provide all information (depending on the model) about the constraints and conditions for examination, as simulation does not give the answers by itself.

Test Yourself 1

Which of the following statements is NOT true in the case of simulation?

A Simulation models are used to model the behaviour of a system
B Simulation models are used to solve complex problems which cannot be solved by analytical techniques
C Simulation models can be used to study alternative solutions to a problem
D Simulation helps to determine the equation describing the situation in linear model

2.	Understand the Monte Carlo simulation model and the method of constructing it. Understand the use of the simulation model for solving queuing problems. Apply simulation techniques to accounting and business situations. [Learning Outcome b,c and d]

2.1 Monte Carlo Simulation Model

Meaning

The Monte Carlo Simulation Model uses random numbers to determine the solution for a complex problem. A complex problem is defined as a problem where either the organisation cannot develop a mathematical formula, nor can perform physical experiment to determine the solution. In such cases Monte Carlo Simulation is beneficial. It is a method of simulation which uses the sampling techniques to address the situation.

Most business activities, plans and processes are too complex for an analytical solution, but you can build a spread sheet model that allows you to evaluate your plan numerically. You can always change numbers and by using different numbers, see the results. This is straightforward if you have just one or two parameters to explore. But in most of the business situations, there are uncertainties in many dimensions, for example, variable market demand, unknown plans of competitors, uncertainty in costs, uncertainty in labour and many others. Therefore, where you are required to make an estimate, forecast or decision where there is a significant amount of uncertainty, the Monte Carlo simulation model shall be of great help as it outlines a method of using random numbers to depict the complicated real life situations and predict results.

Knowledge required for using the Monte Carlo model

To use the Monte Carlo simulation model, you must be able to build a quantitative model of your business activity, plan or process. You also need to know the basics of probability and statistics. To deal with uncertainties in your model, you would need to replace certain fixed numbers with functions that draw random samples from probability distributions. Then, in order to analyze the results of a simulation run, you can use statistics such as mean, standard deviation and percentiles, as well as charts and graphs.

Method of construction of Monte Carlo Simulation

The following steps needs to be followed for the construction of Monte Carlo simulation:

Step 1	Select the measure of effectiveness of the problem
Step 2	Identify the variables which influence the measure of effectiveness significantly
Step 3	Establish a probability distribution of the variables to be analysed
Step 4	Find the cumulative probability distribution for each variable
Step 5	Set Random Number intervals for variables and generate random numbers
Step 6	Simulate the experiment by selecting random numbers from random number tables until the required number of simulations are generated
Step 7	Examine the results and validate the model

2.2 Monte Carlo model and random numbers

Situations in which the technique of simulation is used mostly do not have a specific pattern; i.e. there is randomness involved. For example, the arrival of customers at a grocery shop is random and unpredictable. The “randomness” can be expressed in terms of random numbers. Random numbers are assumed to to imitate a real life situation.

Random numbers are two digit numbers in the range of 00 to 99. Such numbers can be generated by a computer, and are often listed in the published statistics tables.

For example, here is a set of random numbers:

02	14	25	36	48
87	99	03	20	77

In any given situation, every random number has equal possibility of occurrence. There is no pattern, and thus, no way of predicting the appearance of the random numbers, as we can never predict which number will be next in the sequence.

Let us understand the Monte Carlo Simulation Model with the help of an example given below:

Example

A pastry shop's record of the previous month's sale of a particular type of pastry is as follows:

Demand (No. of Pastries)	No. of Days
4	5
5	10
6	6
7	8
8	1

Required:

Simulate the demand for the first 10 days of the month.

Answer

Step 1: Select the measure of effectiveness of the problem

In this case, the measure of effectiveness is the no. of days for which the demand for pastries is to be established.

Step 2: Identify the variables which influence the measure of effectiveness significantly

In this case, the variable which influences the measure of effectiveness is the no. of days.

Step 3: Establish a probability distribution of the variables to be analysed

The probability distribution of demand is determined by expressing the frequencies in terms of proportion ascertained by dividing each value by 30 (as there are on an average 30 days in a month).

Demand (No. of Pastries)	No. of Days	Probability
4	5	0.17
5	10	0.33
6	6	0.20
7	8	0.27
8	1	0.03

Step 4: Find the cumulative probability distribution for each variable

Demand (No. of Pastries)	No. of Days	Probability	Cumulative Probability
4	5	0.17	0.17
5	10	0.33	0.50
6	6	0.20	0.70
7	8	0.27	0.97
8	1	0.03	1.00

Continued on the next page

Step 5: Set random number intervals for variables and generate random numbers

In this case, the probability figures are in two digits, hence we use two digit random numbers taken from a random number table. The random numbers are selected from the table from any row or column, but in a consecutive manner and random intervals are set using the cumulative probability distribution.

Demand (No. of pastries)	No. of Days	Probability	Cumulative Probability	Random Number Interval
4	5	0.17	0.17	00-16
5	10	0.33	0.50	17-49
6	6	0.20	0.70	50-69
7	8	0.27	0.97	70-96
8	1	0.03	1.00	97-99

To simulate the demand for ten days, we select 10 random numbers from the random number table. The random numbers selected are: 17, 46, 85, 09, 50, 58, 04, 77, 69 and 74.

Step 6: Simulate the experiment by selecting random numbers from random number tables until the required number of simulations are generated.

The first random number selected, 17, lies between the random number interval 17-49, corresponding to a demand for 5 pastries per day. Hence, the demand for day one is 5. Similarly, the demand for the remaining days is simulated.

Day	1	2	3	4	5	6	7	8	9	10
Random Number	17	46	85	09	50	58	04	77	69	74
Demand (No. of Pastries)	5	5	7	4	6	6	4	7	6	7

Step 7: Examine the results and validate the model

In this case, the demand for the desired variety of pastry seems to match the actual demand. Therefore, we can take this as the final result. Or the simulation model can be run again with a new set of random numbers and a new result shall be obtained. This process can be repeated till you get a satisfactory result.

Test Yourself 2

The purpose of using random numbers to simulate a situation is:

A To give random outcomes
B To assign values to the parameters
C To ensure that solutions are unbiased
D To describe the uncertainty and randomness of the problem

2.3 Simulation and Queuing problems

The time spent by people in the waiting line, whether at a bank, grocery store or a doctor's clinic, is a valuable resource. Therefore, reduction of the waiting time is an important part of the operations management. Waiting time minimisation is important because quick service is now synonymous with quality service.

A major application of simulation is the analysis of waiting lines or queuing systems. The use of simulation is more suitable for determining a solution to queuing problems as these problems are difficult to put into analytical or mathematical formulae.

Example

Customers arrive at a beauty salon and form a waiting line which feeds a number of tables and attendants. The arrival rate of customers will decide the number of attendants and tables required.

A queuing system can be divided into four components:

1. Arrivals: this is the pattern in which items, peoples and cars arrive in the system.
2. Queue or waiting line: this is about what happens with the people, items and cars between the time of their arrival and the time the service is carried out.
3. Service: this is the time taken to provide the service.
4. Outlet or departure: this is about the exit from the system.

Important

The following are certain key points that need to be understood while solving a queuing problem:

Queue discipline: first come – first served

Service time: it is the length of time taken to serve customers. The service time is mostly random in all situations.

Inter-arrival time: it is the time between the arrival of successive customers in a queuing situation.

Arrival time: arrival time of customers is random in all situations.

Let us understand how to solve a queuing problem with the help of an example.

Example

The arrival time of customers at a bank is according to the following distribution:

Inter-arrival time (IAT) (in minutes)	Probability
3	0.1
4	0.2
5	0.5
6	0.1
7	0.1

Service time (ST) (in minutes)	Probability
3	0.3
4	0.2
5	0.1
6	0.1
7	0.3

Required:

Simulate the process for ten arrivals and estimate the average waiting time for the customer and the percentage idle time for the bank server.

Use the following random numbers:

For IAT : 25, 19, 64, 82, 62, 74, 29, 92, 24, 23, 68, 96
For ST : 92, 41, 66, 07, 44, 29, 52, 43, 87, 55, 47, 83

Assume that the bank opens at 9:00 AM in the morning.

Continued on the next page

Answer

Step 1: Select the measure of effectiveness of the problem

In this case, the measure of effectiveness is the inter-arrival time and the service time which need to be estimated.

Step 2: Identify the variables which influence the measure of effectiveness significantly

In this case, the variables which influence the measure of effectiveness is the inter arrival time and the service time.

Step 3: Establish a probability distribution for the variables to be analysed.

Since the probability distribution of the inter-arrival time and the service time is already provided, we need not calculate it.

Step 4: Find the cumulative probability distribution for each variable

The cumulative probability for both the inter-arrival time and the service time is calculated as shown in the tables given below:

Inter-arrival time (IAT) (in minutes)	Probability	Cumulative probability
3	0.1	0.1
4	0.2	0.3
5	0.5	0.8
6	0.1	0.9
7	0.1	1.0

Service time (ST) (in minutes)	Probability	Cumulative probability
3	0.3	0.3
4	0.2	0.5
5	0.1	0.6
6	0.1	0.7
7	0.3	1.0

Step 5: Set Random Number intervals for variables and generate random numbers

In this case, the random numbers are already provided. We need to set the random number intervals as per the cumulative probability distribution, e.g. for a cumulative probability of 0.1, we allocate random numbers 00 – 0.09.

Inter-arrival time (IAT) (in minutes)	Probability	Cumulative probability	Random Number Intervals
3	0.1	0.1	0.00 – 0.09
4	0.2	0.3	0.10 – 0.29
5	0.5	0.8	0.30 – 0.79
6	0.1	0.9	0.80 – 0.89
7	0.1	1.0	0.90 – 0.99

Service time (ST) (in minutes)	Probability	Cumulative probability	Random Number Intervals
3	0.3	0.3	0.00 – 0.29
4	0.2	0.5	0.30 – 0.49
5	0.1	0.6	0.50 – 0.59
6	0.1	0.7	0.60 – 0.69
7	0.3	1.0	0.70 – 0.99

Continued on the next page

Step 6: Simulate the experiment by selecting random numbers from random number tables until the required number of simulations are generated.

As provided, the following random numbers are to be used:

For IAT: 25, 19, 64, 82, 62, 74, 29, 92, 24, 23, 68, and 96
For ST : 92, 41, 66, 07, 44, 29, 52, 43, 87, 55, 47, 83

Now, we need to run the simulation by allotting the given random numbers to the various inter-arrival and service times according to the selected random number intervals:

Inter-arrival time (IAT) (in minutes)	Random Number Intervals
3	0.00 – 0.09
4	0.10 – 0.29
5	0.30 – 0.79
6	0.80 – 0.89
7	0.90 – 0.99

Service time (ST) (in minutes)	Random Number Intervals
3	0.00 – 0.29
4	0.30 – 0.49
5	0.50 – 0.59
6	0.60 – 0.69
7	0.70 – 0.99

Customer	Random number	Inter-arrival time	Random number	Service time	Time of arrival	Service starts at	Service ends at	Waiting time (mins)	Idle time (mins)
1	25	4	92	7	*9.04	*9.04	9.11	-	4
2	19	4	41	4	9.08	9.11	9.15	3	-
3	64	5	66	6	9.13	9.15	9.21	2	-
4	82	6	07	3	9.19	9.21	9.24	2	-
5	62	5	44	4	9.24	9.24	9.28	-	-
6	74	5	29	3	9.29	9.29	9.32	-	1
7	29	4	52	5	9.33	9.33	9.38	-	1
8	92	7	43	4	9.40	9.40	9.44	-	2
9	24	4	87	7	9.44	9.44	9.51	-	-
10	23	4	55	5	9.48	9.51	9.56	3	-
								10	8

* Note: Since the arrival time of the first customer in the bank is 9:04 am, the start time of the service is considered to be 9:04 am.

Total Time (From 9:00 am to 9:55 am)	56 mins
Total Waiting Time	10 mins
Idle Time	8 mins
Average waiting time per customer	10/10 = 1 minute
Percentage idle time for bank server	(8/56) x 100 = 14.28%

Step 7: Examine the results and validate the model

In this case, since the random numbers are already provided, we do not need to run the simulation again. The given result can be taken as the final result.

If the random numbers are to be selected by us, we need to run the simulation many times, till we get a satisfactory result.

2.4 Simulation and Inventory problems

Managing inventory is one of the most challenging tasks in any kind of business operations. Maintaining the right amount of inventory is very important for any business to reduce its costs as well as to ensure the smooth running of its operations. But in any type of industry, the demand for a certain product cannot be completely predicted, as it depends upon customer preferences, seasonal changes, etc. Moreover, the supply of raw material, labour, machinery etc. may also be affected by numerous factors.

Therefore, in situations where the output of a product and the supply of inputs is difficult to forecast, the simulation technique helps in forecasting and maintaining the inventory. Inventory planning has been discussed in detail in Study Guide 10.

Example

A dealer of commercial bearings has estimated the probability distribution of demand per day for a special type of bearing used in commercial vehicles as follows:

Demand Units	2	3	4	5	6	7	8	9	10
Probability	0.05	0.07	0.09	0.15	0.20	0.21	0.10	0.07	0.06

The various costs involved are:

Ordering Cost	Tshs50,000 per order
Holding Cost	Tshs1,000 per unit per day
Shortage costs	Tshs2,000 per unit per day

A standard lead time of 3 days is required to acquire the inventory.

Random numbers to be used to simulate 20 days of demand units are as follows:

Day	Random number (demand)
1	58
2	45
3	43
4	36
5	46
6	46
7	70
8	32
9	12
10	40
11	51
12	59
13	54
14	16
15	68
16	45
17	96
18	04
19	83
20	77

Required:

Prepare a simulation model a period of 20 days, assuming the reorder quantity is 30 units and re-order level is 20 units, with an opening inventory balance of 45 units and calculate

(i) Re-Ordering costs
(ii) Holding costs
(iii) Shortage costs
(iv) Total costs

Continued on the next page

Answer

Demand per day	Probability	Cumulative probability	Random Number Interval
2	0.05	0.05	00 - 04
3	0.07	0.12	05 - 11
4	0.09	0.21	12 - 20
5	0.15	0.36	21 - 35
6	0.20	0.56	36 - 55
7	0.21	0.77	56 - 76
8	0.10	0.87	77 - 86
9	0.07	0.94	87 - 93
10	0.06	1.00	94 - 99

Reorder Quantity = 30 units, Reorder Level = 20 units, Beginning Inventory = 45 units

Lead time is the time between placement of order and the delivery of the same

Day	Random Number (Demand)	Demand	Lead Time (Days)	Inventory at end of day	Quantity Received	Ordering Cost	Holding Cost
		Units	Days	Units	Units	Tshs'000	Tshs'000
0	-	-	-	45	-	-	
1	58	7	-	38	-	-	38
2	45	6	-	32	-	-	32
3	43	6	-	26	-	-	26
4	36	6	3	20	-	50	20
5	46	6	-	14	-	-	14
6	46	6	-	8	-	-	8
7	70	7	-	31	30	-	31
8	32	5	-	29	-	-	29
9	12	4	-	25	-	-	25
10	40	6	-	19	-	50	19
11	51	6	2	13	-	-	13
12	59	7	-	6	-	-	6
13	54	6	-	30	30	-	30
14	16	4	-	26	-	-	26
15	68	7	-	19	-	50	19
16	45	6	2	13	-	-	13
17	96	10	-	3	-	-	3
18	04	2	-	31	30	-	31
19	83	8	-	23	-	-	23
20	77	8	-	15	-	50	15
					90	200	421

The above table shows a simulation plan for 20 days, which calls for a reorder quantity of 30 units and a re-order level of 20 units. Further the table shows the calculation of demand, inventory level, quantity received, ordering cost and holding cost

Continued on the next page

Based on the simulation done above, the various costs can be calculated as follows:

(i) Total re-ordering cost = 4 orders X Tshs50,000 = Tshs200,000

(ii) Total holding cost = Tshs421,000 (refer table above)

(iii) Since the demand for each day is satisfied, there is no shortage cost.

(iv) Total Cost = Tshs200,000 + Tshs421,000 = Tshs621,000

Test Yourself 3

Ramco Sales, a dealer in electronic appliances, sells a particular model of TV, for which the probability distribution of daily demand is as given in the table below:

Demand/day	0	1	2	3	4	5
Probability	0.05	0.25	0.20	0.25	0.10	0.15

Use the following random numbers to simulate the demand for the next ten days and find out the average demand for TV sets per day:

68, 47, 92, 76, 86, 46, 16, 28, 35, 54.

A 3.0
B 3.5
C 2.8
D 1.5

Test Yourself 4

A farmer has 20 acres of land and cultivates rice on the entire land. Due to fluctuations in the rain water availability, the yield per acre differs. The probability distribution yields are given below:

Yield of rice per acre	Probability
200	0.15
220	0.25
240	0.35
260	0.13
280	0.12

The farmer wants to know the average yield for the next 12 months if the same water availability continues.

Simulate the average yield using the following random numbers:
50, 28, 68, 36, 90, 62, 27, 50, 18, 36, 61 and 21.

A 233.3
B 250.5
C 299.8
D 190.5

Answers to Test Yourself

Answer to TY 1

The correct option is D.

Answer to TY 2

The correct option is D.

The purpose of random numbers is to describe uncertainty and randomness of the problem.

Answer to TY 3

The correct option is C.

Table for Random Number Interval for demand

Demand	Probability	Cumulative Probability	Random Number Intervals
0	0.05	0.05	00 – 04
1	0.25	0.30	05 – 29
2	0.20	0.50	30 – 49
3	0.25	0.75	50 – 74
4	0.10	0.85	75 – 84
5	0.15	1.00	85 – 99

Expected demand as per ten Random Numbers selected

Trial No.	Random Number	Demand/ day
1	68	3
2	47	2
3	92	5
4	76	4
5	86	5
6	46	2
7	16	1
8	28	1
9	35	2
10	54	3
Total Demand		28
Average Demand / per day		28/10 = 2.8

Answer to TY 4

The correct option is A.

Table for Random Number Interval for yield

Yield of rice per acre	Probability	Cumulative Probability	Random Number Interval
200	0.15	0.15	00 – 14
220	0.25	0.40	15 – 39
240	0.35	0.75	40 – 74
260	0.13	0.88	75 – 87
280	0.12	1.00	88 – 99

Table for Simulation for a period of twelve months:

Month	Random Number	Random Number Interval	Expected Yield of rice per acre
1	50	40 – 74	240
2	28	15 – 39	220
3	68	40 – 74	240
4	36	15 – 39	220
5	90	88 – 99	280
6	62	40 – 74	240
7	27	15 – 39	220
8	50	40 – 74	240
9	18	15 – 39	220
10	36	15 – 39	220
11	61	40 – 74	240
12	21	15 – 39	220
		Total Yield for the year	2,800
		Average yield for the year	2,800/12 = 233.33

Self Examination Questions

Question 1

Explain the concept of simulation and describe the steps in the simulation process.

Question 2

Which one of the following options is not a step involved in the Monte Carlo method?

A Identifying variables which influence the experiment
B Specifying the numbers on the roulette wheel
C Establishing probability distribution of the variables
D Simulating the experiment using random numbers

Question 3

A candy shop has observed that the demand for their strawberry candy packets per week is as follows:

Demand / Week	5	10	15	20	25	30	35
Frequency	4	10	35	25	15	9	2

Random numbers - 68, 47, 92, 76, 86, 46, 16, 28, 35, 54

Required:

Generate the demand for the next 10 weeks, and also find the average demand.

Question 4

A company manufactures a component which requires a high degree of precision. Each unit of the component is therefore subjected to a strict quality control test to ascertain whether there is any defect in it. The defects are classified into three categories viz. A, B and C. If defect A occurs in any item, it is scrapped. If defect B or C occurs in any item, it is reworked upon to rectify the defect.

The machine time involved in rework for defect B in any product is 30 minutes and that for defect C is 45 minutes.

The probabilities are as under:

	Defect A	Defect B	Defect C
Defect occurring	0.15	0.20	0.10
Defect not occurring	0.85	0.80	0.90

Random numbers for defect A	48	55	91	40	93	01	83	63	47	52
Random numbers for defect B	47	36	57	04	79	55	10	13	57	09
Random numbers for defect C	82	95	18	96	20	84	56	11	52	03

Required:

Using the above random numbers, simulate a study of 10 items of output and determine the following:

(i) the number of items with no defects
(ii) the number of items scrapped due to occurrence of defect A
(iii) the total machine time required for rework due to occurrence of defect B or C

Question 5

Ben started a Hot Dog shop. He entered into a contract with a bakery to buy readymade hot dogs at a cost of Tshs18,000 each and sell them for Tshs25,000 each. He estimates the sales pattern of the hot dog as follows:

Demand (No. of copies)	Probability
0 < 100	0.18
100 < 200	0.32
200 < 350	0.25
350 < 450	0.15
450 < 500	0.06
500 < 600	0.04

Ben has contracted for 250 hot dogs per month from the bakery. The unsold hot dogs are disposed of at half the price each.

Ben wants to simulate the demand and profitability. The following random numbers may be used for simulation:

27, 15, 56, 17, 98, 71, 51, 32, 62, 83, 96, 69.

Required:

(i) Allocate random numbers to the demand pattern forecast by Ben.
(ii) Simulate twelve months' sales and calculate the monthly and annual profit/loss.
(iii) Calculate the loss on lost sales

Answers to Self Examination Questions

Answer to SEQ 1

Simulation is a technique to determine the most likely result of a real life situations for which a mathematical model or a real life experiment is not possible or practicable. Simulation models are used to simulate large and complex problems. The simulation is generally performed using the Monte Carlo Method of Simulation and involves the use of random numbers.

Steps in Simulation

The following are the steps that are usually followed in a simulation process:

Step1: The first step is to define the problem

Step 2: The variables of the model that have a logical relationship with one another are introduced and a suitable model is constructed.

Step 3: After developing a desired model, each alternative is evaluated by generating a series of values of the random variable, and the behaviour of the system is observed.

Step 4: Finally, the results are examined and the best alternative is selected.

Answer to SEQ 2

The correct option is B.

Answer to SEQ 3

Table for Random Number Interval for demand

Demand	Probability	Cumulative Probability	Random Number Intervals
5	0.04	0.04	00 – 03
10	0.10	0.14	04 – 13
15	0.35	0.49	14 – 48
20	0.25	0.74	49 – 73
25	0.15	0.89	74 – 88
30	0.09	0.98	89 – 97
35	0.02	1.00	98 - 99

Expected Demand as per ten Random Numbers selected

Trial No.	Random Number	Demand/ week
1	68	20
2	47	15
3	92	35
4	76	25
5	86	25
6	46	15
7	16	15
8	28	15
9	35	15
10	54	20
Total Demand		200
Average Demand / week		200/10 = 20

Answer to SEQ 4

Defect A		Defect B		Defect C	
Defect Exists or not	Random Number Allocation	Defect Exists or not	Random Number Allocation	Defect Exists or not	Random Number Allocation
Yes	00 –14	Yes	00 – 19	Yes	00 – 09
No	15 – 99	No	20 – 99	No	10 – 99

Simulation:

Item No	Random No.			Whether defect exists?	Items scrapped	Rework Minutes
	Defect A	Defect B	Defect C			
1	48	47	82	No	-	-
2	55	36	95	No	-	-
3	91	57	18	No	-	-
4	40	04	96	B		30
5	93	79	20	No		
6	01	55	84	A	1	-
7	83	10	56	B	-	30
8	63	13	11	B	-	30
9	47	57	52	No		
10	52	09	03	B and C	-	75
Total					1	165

No. of items without any defect	5
No. of items with defect A	1
No. of items scrapped	1
No. of items with defect C	1
No. of items with defect B	4
Rework minutes to rectify defect B	120 minutes
Rework minutes to rectify defect C	45 minutes
Total rework minutes	165 Minutes

Answer to SEQ 5

(i) Allocation of random numbers

Demand	Probability	Cumulative Probability	Allocation of Random Numbers
00 < 100	0.18	0.18	00 – 17
100 < 200	0.32	0.50	17 – 49
200 < 350	0.25	0.75	49 – 74
350 < 450	0.15	0.90	74 – 89
450 < 500	0.06	0.96	89 – 96
500 < 600	0.04	1.00	96 – 99

(ii) Simulation: twelve months' sales, monthly and annual profit/loss

Tshs25,000 – Tshs18.,000

Loss on unsold hotdogs = 50% of cost = Tshs x 50% = Tshs9,000

Month	Random Numbers	Hot Dog Demand	Hot Dogs Sold	Hot Dogs Return	Profit on sales (Hot Dogs sold X Tshs7,000))	Loss on disposal (Hot Dogs disposed X Tshs9,000)	Net Profit/ Loss	Loss on units falling short of demand (excess demand x Tshs7,000)
		Units	Units	Units	Tshs'000	Tshs'000	Tshs'000	Tshs'000
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(f – g)	(c – d) x Tshs7,000
1	27	150	150	100	1,050	900	150	0
2	15	50	50	200	350	1,800	(1,450)	0
3	56	275	250	0	1,750	0	1,750	175
4	17	50	50	200	350	1,800	(1,450)	0
5	98	550	250	0	1,750	0	1,750	2,100
6	71	275	250	0	1,750	0	1,750	175
7	51	275	250	0	1,750	0	1,750	175
8	32	150	150	100	1,050	900	150	0
9	62	275	250	0	1,750	0	1,750	175
10	83	400	250	0	1,750	0	1,750	1,050
11	96	550	250	0	1,750	0	1,750	2,100
12	69	275	250	0	1,750	0	1,750	175
				Total	16,800	5,400	11,400	6,125

Profit on the hot dogs sold = Tshs16,800,000

Loss on disposal of un-sold hot dogs = Tshs5,400,000

Net profit during the year = Tshs16,800,000 – Tshs5,400,000 = Tshs11,400,000

Net profit per month = Tshs11,400,000/12 months = Tshs950,000

(iii) Loss due to lost sales = Tshs6,125,000 (refer table above for calculation)

APPENDICES

Area under standard normal curve

	0	0_01	0_02	0.03	0.04	0.05	0_06	0.07	0.08	0.09
0	0	0_004	0.008	0.012	0.016	0.0199	0.0239	0.0279	0.0319	0.0359
0_1	0.0398	0_0438	0_0478	0.0517	0.0557	0.0596	0_0636	0.0675	0-0714	0.0753
0_2	0.0793	0_0832	0-0871	0-091	0.0948	0.0987	0_1026	0.1064	0.1103	0.1141
0.3	0.1179	0_1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.148	0.1517
0_4	0.1554	0_1591	0.1628	0.1664	0.17	0.1736	0.1772	0.1808	0.1844	0.1879
0_5	0.1915	0_195	0_1985	0.2019	0.2054	0.2088	0_2123	0.2157	0.219	0.2224
0_6	0.2257	0-2291	0.2324	0.2357	0.2389	0.2422	0_2454	0.2486	0.2517	0.2549
0.7	0.258	0-2611	0.2642	0.2673	0.2704	0-2734	0.2764	0.2794	0.2823	0.2852
0_8	0.2881	0-291	0.2939	0.2967	0.2995	0.3023	0_3051	0.3078	0.3106	0.3133
0_9	0.3159	0_3186	0_3212	0.3238	0.3264	0.3289	0_3315	0.334	0.3365	0.3389
1	0-3413	0-3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1_1	0.3643	0_3665	0.3686	0.3708	0.3729	0-3749	0-377	0.379	0.381	0.383
1_2	0.3849	0_3869	0.3888	0.3907	0-3925	0-3944	0-3962	0.398	0.3997	0.4015
1_3	0_4032	0.4049	0_4066	0_4082	0_4099	0- 4115	0- 4131	0- 4147	0- 4162	0_4177
1_4	0_4192	0.4207	0_4222	0_4236	0_4251	0_4265	0_4279	0- 4292	0- 4306	0- 4319
1_5	0- 4332	0- 4345	0_4357	0_437	0_4382	0_4394	0- 4406	0- 4418	0- 4429	0-4441
1_6	0- 4452	0- 4463	0_4474	0- 4484	0- 4495	0_4505	0_4515	0_4525	0_4535	0_4545
1_7	0_4554	0.4564	0_4573	0_4582	0- 4591	0_4599	0- 4608	0_4616	0_4625	0_4633
1_8	0-4641	0.4649	0_4656	0_4664	0_4671	0_4678	0_4686	0_4693	0_4699	0_4706
1_9	0- 4713	0.4719	0_4726	0_4732	0_4738	0_4744	0_475	0_4756	0_4761	0_4767
2	0- 4772	0.4778	0-4783	0-4788	0-4793	0-4798	0_4803	0_4808	0_4812	0_4817
2.1	0_4821	0_4826	0.483	0-4834	0.4838	0_4842	0.4846	0_485	0.4854	0_4857
2.2	0_4861	0_4864	0_4868	0_4871	0_4875	0_4878	0_4881	0_4884	0_4887	0_489
2.3	0_4893	0_4896	0_4898	0.4901	0.4904	0.4906	0_4909	0-4911	0-4913	0-4916
2.4	0- 4918	0-492	0-4922	0-4925	0-4927	0-4929	0-4931	0.4932	0.4934	0.4936
2.5	0_4938	0_494	0-4941	0.4943	0.4945	0.4946	0_4948	0.4949	0.4951	0.4952
2.6	0_4953	0_4855	0_4956	0_4957	0_4959	0-496	0- 4961	0- 4962	0- 4963	0- 4964
2_7	0_4965	0_4966	0_4967	0.4968	0.4969	0.497	0-4971	0-4972	0.4973	0.4974
2_8	0_4974	0_4975	0_4976	0.4977	0.4977	0.4978	0_4979	0.4979	0.498	0_4981
2_9	0_4981	0_4982	0.4982	0.4983	0.4984	0.4984	0_4985	0.4985	0.4986	0.4986
3	0_4987	0_4987	0_4987	0.4988	0.4988	0.4989	0_4989	0.4989	0.499	0.499

Student t-Table

Alpha	0.250	0.200	0.150	0.100	0.050	0.025	0.010	0.005	0.0005
df									
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.656	636.578
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.600
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.307	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1 093	1372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1 083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1 074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1 066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850

Present Value Table

Present value of 1 i.e. $(1 + r)^{-n}$

Where r = discount rate

n = number of periods until payment

Discount rate (r)

Periods (n)	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	
1	0·990	0·980	0·971	0·962	0·952	0·943	0·935	0·926	0·917	0·909	1
2	0·980	0·961	0·943	0·925	0·907	0·890	0·873	0·857	0·842	0·826	2
3	0·971	0·942	0·915	0·889	0·864	0·840	0·816	0·794	0·772	0·751	3
4	0·961	0·924	0·888	0·855	0·823	0·792	0·763	0·735	0·708	0·683	4
5	0·951	0·906	0·863	0·822	0·784	0·747	0·713	0·681	0·650	0·621	5
6	0·942	0·888	0·837	0·790	0·746	0·705	0·666	0·630	0·596	0·564	6
7	0·933	0·871	0·813	0·760	0·711	0·665	0·623	0·583	0·547	0·513	7
8	0·923	0·853	0·789	0·731	0·677	0·627	0·582	0·540	0·502	0·467	8
9	0·914	0·837	0·766	0·703	0·645	0·592	0·544	0·500	0·460	0·424	9
10	0·905	0·820	0·744	0·676	0·614	0·558	0·508	0-463	0-422	0·386	10
11	0·896	0·804	0·722	0·650	0·585	0·527	0·475	0·429	0·388	0·350	11
12	0·887	0·788	0·701	0·625	0·557	0·497	0·444	0·397	0·356	0·319	12
13	0·879	0·773	0·681	0·601	0·530	0·469	0-415	0·368	0·326	0·290	13
14	0·870	0·758	0·661	0·577	0·505	0·442	0·388	0·340	0·299	0·263	14
15	0·861	0·743	0·642	0·555	0·481	0·417	0·362	0·315	0·275	0·239	15

(n)	11%	12%	13%	14%	15%	16%	17%	18%	19%	20%	
1	0·901	0·893	0·885	0·877	0·870	0·862	0·855	0·847	0·840	0·833	1
2	0·812	0·797	0·783	0·769	0·756	0·743	0·731	0·718	0·706	0·694	2
3	0·731	0·712	0·693	0·675	0·658	0·641	0·624	0·609	0·593	0·579	3
4	0·659	0·636	0·613	0·592	0·572	0·552	0·534	0·516	0·499	0·482	4
5	0·593	0·567	0·543	0·519	0·497	0·476	0·456	0·437	0·419	0·402	5
6	0·535	0·507	0·480	0·456	0·432	0·410	0·390	0·370	0·352	0·335	6
7	0·482	0·452	0·425	0·400	0·376	0·354	0·333	0·314	0·296	0·279	7
8	0·434	0·404	0·376	0·351	0·327	0·305	0·285	0·266	0·249	0·233	8
9	0·391	0·361	0·333	0·308	0·284	0·263	0·243	0·225	0·209	0·194	9
10	0·352	0·322	0·295	0·270	0·247	0·227	0·208	0·191	0·176	0·162	10
11	0·317	0·287	0·261	0·237	0·215	0·195	0·178	0·162	0·148	0·135	11
12	0·286	0·257	0·231	0·208	0·187	0·168	0·152	0·137	0·124	0·112	12
13	0·258	0·229	0·204	0·182	0·163	0·145	0·130	0·116	0·104	0·093	13
14	0·232	0·205	0·181	0·160	0·141	0·125	0·111	0·099	0·088	0·078	14
15	0·209	0·183	0·160	0·140	0·123	0·108	0·095	0·084	0·074	0·065	15

Annuity Table

Present value of an annuity of 1 i.e. $\frac{1-(1+r)^{-n}}{r}$

Where r = discount rate
n = number of periods

Discount rate (r)

Periods (nl	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	
1	0·990	0·980	0·971	0·962	0·952	0·943	0·935	0·926	0·917	0·909	1
2	1·970	1·942	1·913	1·886	1·859	1·833	1-808	1·783	1·759	1·736	2
3	2·941	2·884	2·829	2·775	2·723	2·673	2·624	2·577	2·531	2-487	3
4	3·902	3·808	3·717	3·630	3·546	3-465	3·387	3·312	3·240	3·170	4
5	4·853	4·713	4·580	4·452	4·329	4·212	4·100	3·993	3·890	3·791	5
6	5·795	5·601	5·417	5·242	5·076	4·917	4·767	4·623	4·486	4·355	6
7	6·728	6·472	6·230	6·002	5·786	5·582	5·389	5·206	5·033	4·868	7
8	7·652	7·325	7·020	6·733	6-463	6·210	5·971	5·747	5·535	5·335	8
9	8·566	8·162	7·786	7·435	7-108	6·802	6·515	6·247	5·995	5·759	9
10	9·471	8·983	8·530	8·111	7-722	7·360	7·024	6·710	6-418	6·145	10
11	10·37	9·787	9·253	8·760	8·306	7·887	7-499	7·139	6·805	6-495	11
12	11·26	10·58	9·954	9·385	8·863	8·384	7·943	7·536	7·161	6·814	12
13	12·13	11·35	10·63	9·986	9·394	8·853	8·358	7·904	7·487	7·103	13
14	13·00	12·11	11·30	10·56	9·899	9·295	8·745	8·244	7·786	7·367	14
15	13·87	12·85	11·94	11·12	10·38	9·712	9·108	8·559	8·061	7·606	15

(nl	11%	12%	13%	14%	15%	16%	17%	18%	19%	20%	
1	0·901	0·893	0·885	0·877	0·870	0·862	0·855	0·847	0·840	0·833	1
2	1·713	1·690	1·668	1·647	1·626	1·605	1·585	1·566	1·547	1·528	2
3	2-444	2-402	2·361	2·322	2·283	2·246	2·210	2·174	2·140	2·106	3
4	3·102	3·037	2·974	2·914	2·855	2·798	2·743	2·690	2·639	2·589	4
5	3·696	3·605	3·517	3·433	3·352	3·274	3·199	3·127	3·058	2·991	5
6	4·231	4·111	3·998	3·889	3·784	3·685	3·589	3-498	3-410	3·326	6
7	4·712	4·564	4·423	4·288	4·160	4·039	3·922	3·812	3·706	3·605	7
8	5·146	4·968	4·799	4·639	4-487	4·344	4·207	4·078	3·954	3·837	8
9	5·537	5·328	5·132	4·946	4·772	4·607	4-451	4·303	4·163	4·031	9
10	5·889	5·650	5·426	5·216	5·019	4·833	4·659	4·494	4·339	4·192	10
11	6·207	5·938	5·687	5-453	5·234	5·029	4·836	4·656	4·486	4·327	11
12	6·492	6·194	5·918	5·660	5·421	5·197	4·988	4·793	4·611	4·439	12
13	6·750	6·424	6·122	5·842	5·583	5·342	5·118	4·910	4·715	4·533	13
14	6·982	6·628	6·302	6·002	5·724	5-468	5·229	5·008	4·802	4·611	14
15	7-191	6·811	6·462	6·142	5·847	5·575	5·324	5·092	4·876	4·675	15

Critical values

	p value											
df	0.25	0.2	0.15	0.1	0.05	0.025	0.02	0.01	0.005	0.003	0.001	0.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.59	6.74	7.78	9.49	11.14	11.67	13.23	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.33	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.53	14.45	15.03	16.81	13.55	20.25	22.46	24.10
7	9.04	5.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.63	21.67	23.59	25.46	27.83	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42
11	13.70	14.63	15.77	17.29	19.68	21.92	22.62	24.72	26.76	28.73	31.26	33.14
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91	34.82
13	15.93	15.58	18.90	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53	36.48
14	17.12	18.15	19.40	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12	38.11
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70	39.72
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25	41.31
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79	42.88
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31	44.43
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82	45.97
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31	47.50
21	24.93	26.17	27.66	29.62	39.67	35.48	36.34	38.93	41.40	43.78	46.80	49.01
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27	50.51
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73	52.00
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18	53.48
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62	54.95
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05	56.41
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48	57.86
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89	59.30
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30	60.73
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70	62.16
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40	76.09
50	56.33	53.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66	89.56
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61	102.70
80	88.13	90.41	93.11	96.58	101.90	106.60	108.10	112.30	116.30	120.10	124.80	128.30
100	109.10	111.70	114.70	118.50	124.30	129.60	131.10	135.80	140.20	144.30	149.40	153.20

Logarithmic Table

LOGARITHIMIC TABLE

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
1.0	0.0000	0.0043	0.0086	0.0128	0.0170	0.0212	0.0253	0.0294	0.0334	0.0374	4	8	12	17	21	25	29	33	37
1.1	0.0414	0.0453	0.0492	0.0531	0.0569	0.0607	0.0645	0.0682	0.0719	0.0755	4	8	11	15	19	23	27	30	34
1.2	0.0792	0.0828	0.0864	0.0899	0.0934	0.0969	0.1004	0.1038	0.1072	0.1106	3	7	10	14	17	21	24	28	31
1.3	0.1139	0.1173	0.1206	0.1239	0.1271	0.1303	0.1335	0.1367	0.1399	0.1430	3	6	10	13	16	19	23	26	29
1.4	0.1461	0.1492	0.1523	0.1553	0.1584	0.1614	0.1644	0.1673	0.1703	0.1732	3	6	9	12	15	18	21	24	27
1.5	0.1761	0.1790	0.1818	0.1847	0.1875	0.1903	0.1931	0.1959	0.1987	0.2014	3	6	8	11	14	17	20	22	25
1.6	0.2041	0.2068	0.2095	0.2122	0.2148	0.2175	0.2201	0.2227	0.2253	0.2279	3	5	8	11	13	16	18	21	24
1.7	0.2304	0.2330	0.2355	0.2380	0.2405	0.2430	0.2455	0.2480	0.2504	0.2529	2	5	7	10	12	15	17	20	22
1.8	0.2553	0.2577	0.2601	0.2625	0.2648	0.2672	0.2695	0.2718	0.2742	0.2765	2	5	7	9	12	14	16	19	21
1.9	0.2788	0.2810	0.2833	0.2856	0.2878	0.2900	0.2923	0.2945	0.2967	0.2989	2	4	7	9	11	13	16	18	20
2.0	0.3010	0.3032	0.3054	0.3075	0.3096	0.3118	0.3139	0.3160	0.3181	0.3201	2	4	6	8	11	13	15	17	19
2.1	0.3222	0.3243	0.3263	0.3284	0.3304	0.3324	0.3345	0.3365	0.3385	0.3404	2	4	6	8	10	12	14	16	18
2.2	0.3424	0.3444	0.3464	0.3483	0.3502	0.3522	0.3541	0.3560	0.3579	0.3598	2	4	6	8	10	12	14	15	17
2.3	0.3617	0.3636	0.3655	0.3674	0.3692	0.3711	0.3729	0.3747	0.3766	0.3784	2	4	6	7	9	11	13	15	17
2.4	0.3802	0.3820	0.3838	0.3856	0.3874	0.3892	0.3909	0.3927	0.3945	0.3962	2	4	5	7	9	11	12	14	16
2.5	0.3979	0.3997	0.4014	0.4031	0.4048	0.4065	0.4082	0.4099	0.4116	0.4133	2	3	5	7	9	10	12	14	15
2.6	0.4150	0.4166	0.4183	0.4200	0.4216	0.4232	0.4249	0.4265	0.4281	0.4298	2	3	5	7	8	10	11	13	15
2.7	0.4314	0.4330	0.4346	0.4362	0.4378	0.4393	0.4409	0.4425	0.4440	0.4456	2	3	5	6	8	9	11	13	14
2.8	0.4472	0.4487	0.4502	0.4518	0.4533	0.4548	0.4564	0.4579	0.4594	0.4609	2	3	5	6	8	9	11	12	14
2.9	0.4624	0.4639	0.4654	0.4669	0.4683	0.4698	0.4713	0.4728	0.4742	0.4757	1	3	4	6	7	9	10	12	13
3.0	0.4771	0.4786	0.4800	0.4814	0.4829	0.4843	0.4857	0.4871	0.4886	0.4900	1	3	4	6	7	9	10	11	13
3.1	0.4914	0.4928	0.4942	0.4955	0.4969	0.4983	0.4997	0.5011	0.5024	0.5038	1	3	4	6	7	8	10	11	12
3.2	0.5051	0.5065	0.5079	0.5092	0.5105	0.5119	0.5132	0.5145	0.5159	0.5172	1	3	4	5	7	8	9	11	12
3.3	0.5185	0.5198	0.5211	0.5224	0.5237	0.5250	0.5263	0.5276	0.5289	0.5302	1	3	4	5	6	8	9	10	12
3.4	0.5315	0.5328	0.5340	0. 353	0.5366	0.5378	0.5391	0.5403	0.5416	0.5428	1	3	4	5	6	8	9	10	11
3.5	0.5441	0.5453	0.5465	0.5478	0.5490	0.5502	0.5514	0.5527	0.5539	0.5551	1	2	4	5	6	7	9	10	11
3.6	0.5563	0.5575	0.5587	0.5599	0.5611	0.5623	0.5635	0.5647	0.5658	0.5670	1	2	4	5	6	7	8	10	11
3.7	0.5682	0.5694	0.5705	0.5717	0.5729	0.5740	0.5752	0.5763	0.5775	0.5786	1	2	3	5	6	7	8	9	10
3.8	0.5798	0.5809	0.5821	0.5832	0.5843	0.5855	0.5866	0.5877	0.5888	0.5899	1	2	3	5	6	7	8	9	10
3.9	0.5911	0.5922	0.5933	0.5944	0.5955	0.5966	0.5977	0.5988	0.5999	0.6010	1	2	3	4	6	7	8	9	10
4.0	0.6021	0.6031	0.6042	0.6053	0.6064	0.6075	0.6085	0.6096	0.6107	0.6117	1	2	3	4	5	6	8	9	10
4.1	0.6128	0.6138	0.6149	0.6160	0.6170	0.6180	0.6191	0.6201	0.6212	0.6222	1	2	3	4	5	6	7	8	9
4.2	0.6232	0.6243	0.6253	0.6263	0.6274	0.6284	0.6294	0.6304	0.6314	0.6325	1	2	3	4	5	6	7	8	9
4.3	0.6335	0.6345	0.6355	0.6365	0.6375	0.6385	0.6395	0.6405	0.6415	0.6425	1	2	3	4	5	6	7	8	9
4.4	0.6435	0.6444	0.6454	0.6464	0.6474	0.6484	0.6493	0.6503	0.6513	0.6522	1	2	3	4	5	6	7	8	9
4.5	0.6532	0.6542	0.6551	0.6561	0.6571	0.6580	0.6590	0.6599	0.6609	0.6618	1	2	3	4	5	6	7	8	9
4.6	0.6628	0.6637	0.6646	0.6656	0.6665	0.6675	0.6684	0.6693	0.6702	0.6712	1	2	3	4	5	6	7	7	8
4.7	0.6721	0.6730	0.6739	0.6749	0.6758	0.6767	0.6776	0.6785	0.6794	0.6803	1	2	3	4	5	5	6	7	8
4.8	0.6812	0.6821	0.6830	0.6839	0.6848	0.6857	0.6866	0.6875	0.6884	0.6893	1	2	3	4	4	5	6	7	8
4.9	0.6902	0.6911	0.6920	0.6928	0.6937	0.6946	0.6955	0.6964	0.6972	0.6981	1	2	3	4	4	5	6	7	8
5.0	0.6990	0.6998	0.7007	0.7016	0.7024	0.7033	0.7042	0.7050	0.7059	0.7067	1	2	3	3	4	5	6	7	8
5.1	0.7076	0.7084	0.7093	0.7101	0.7110	0.7118	0.7126	0.7135	0.7143	0.7152	1	2	3	3	4	5	6	7	8
5.2	0.7160	0.7168	0.7177	0.7185	0.7193	0.7202	0.7210	0.7218	0.7226	0.7235	1	2	2	3	4	5	6	7	7
5.3	0.7243	0.7251	0.7259	0.7267	0.7275	0.7284	0.7292	0.7300	0.7308	0.7316	1	2	2	3	4	5	6	7	7
5.4	0.7324	0.7332	0.7340	0.7348	0.7356	0.7364	0.7372	0.7380	0.7388	0.7396	1	2	2	3	4	5	6	6	7

LOGARITHIMIC TABLE

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
5.5	0.7404	0.7412	0.7419	0.7427	0.7435	0.7443	0.7451	0.7459	0.7466	0.7474	1	2	2	3	4	5	5	6	7
5.6	0.7482	0.7490	0.7497	0.7505	0.7513	0.7520	0.7528	0.7536	0.7543	0.7551	1	2	2	3	4	5	5	6	7
5.7	0.7559	0.7566	0.7574	0.7582	0.7589	0.7597	0.7604	0.7612	0.7619	0.7627	1	2	2	3	4	5	5	6	7
5.8	0.7634	0.7642	0.7649	0.7657	0.7664	0.7672	0.7679	0.7686	0.7694	0.7701	1	1	2	3	4	4	5	6	7
5.9	0.7709	0.7716	0.7723	0.7731	0.7738	0.7745	0.7752	0.7760	0.7767	0.7774	1	1	2	3	4	4	5	6	7
6.0	0.7782	0.7789	0.7796	0.7803	0.7810	0.7818	0.7825	0.7832	0.7839	0.7846	1	1	2	3	4	4	5	6	6
6.1	0.7853	0.7860	0.7868	0.7875	0.7882	0.7889	0.7896	0.7903	0.7910	0.7917	1	1	2	3	4	4	5	6	6
6.2	0.7924	0.7931	0.7938	0.7945	0.7952	0.7959	0.7966	0.7973	0.7980	0.7987	1	1	2	3	3	4	5	6	6
6.3	0.7993	0.8000	0.8007	0.8014	0.8021	0.8028	0.8035	0.8041	0.8048	0.8055	1	1	2	3	3	4	5	5	6
6.4	0.8062	0.8069	0.8075	0.8082	0.8089	0.8096	0.8102	0.8109	0.8116	0.8122	1	1	2	3	3	4	5	5	6
6.5	0.8129	0.8136	0.8142	0.8149	0.8156	0.8162	0.8169	0.8176	0.8182	0.8189	1	1	2	3	3	4	5	5	6
6.6	0.8195	0.8202	0.8209	0.8215	0.8222	0.8228	0.8235	0.8241	0.8248	0.8254	1	1	2	3	3	4	5	5	6
6.7	0.8261	0.8267	0.8274	0.8280	0.8287	0.8293	0.8299	0.8306	0.8312	0.8319	1	1	2	3	3	4	5	5	6
6.8	0.8325	0.8331	0.8338	0.8344	0.8351	0.8357	0.8363	0.8370	0.8376	0.8382	1	1	2	3	3	4	4	5	6
6.9	0.8388	0.8395	0.8401	0.8407	0.8414	0.8420	0.8426	0.8432	0.8439	0.8445	1	1	2	3	3	4	4	5	6
7.0	0.8451	0.8457	0.8463	0.8470	0.8476	0.8482	0.8488	0.8494	0.8500	0.8506	1	1	2	2	3	4	4	5	6
7.1	0.8513	0.8519	0.8525	0.8531	0.8537	0.8543	0.8549	0.8555	0.8561	0.8567	1	1	2	2	3	4	4	5	5
7.2	0.8573	0.8579	0.8585	0.8591	0.8597	0.8603	0.8609	0.8615	0.8621	0.8627	1	1	2	2	3	4	4	5	5
7.3	0.8633	0.8639	0.8645	0.8651	0.8657	0.8663	0.8669	0.8675	0.8681	0.8686	1	1	2	2	3	4	4	5	5
7.4	0.8692	0.8698	0.8704	0.8710	0.8716	0.8722	0.8727	0.8733	0.8739	0.8745	1	1	2	2	3	4	4	5	5
7.5	0.8751	0.8756	0.8762	0.8768	0.8774	0.8779	0.8785	0.8791	0.8797	0.8802	1	1	2	2	3	3	4	5	5
7.6	0.8808	0.8814	0.8820	0.8825	0.8831	0.8837	0.8842	0.8848	0.8854	0.8859	1	1	2	2	3	3	4	5	5
7.7	0.8865	0.8871	0.8876	0.8882	0.8887	0.8893	0.8899	0.8904	0.8910	0.8915	1	1	2	2	3	3	4	4	5
7.8	0.8921	0.8927	0.8932	0.8938	0.8943	0.8949	0.8954	0.8960	0.8965	0.8971	1	1	2	2	3	3	4	4	5
7.	0.8976	0.	0.8987	0.	0.8998	0.	0.9009	0.	0.9020	0.	1		2		3		4		5
8.0	0.9031	0.9036	0.9042	0.9047	0.9053	0.9058	0.9063	0.9069	0.9074	0.9079	1	1	2	2	3	3	4	4	5
8.1	0.9085	0.9090	0.9096	0.9101	0.9106	0.9112	0.9117	0.9122	0.9128	0.9133	1	1	2	2	3	3	4	4	5
8.2	0.9138	0.9143	0.9149	0.9154	0.9159	0.9165	0.9170	0.9175	0.9180	0.9186	1	1	2	2	3	3	4	4	5
8.3	0.9191	0.9196	0.9201	0.9206	0.9212	0.9217	0.9222	0.9227	0.9232	0.9238	1	1	2	2	3	3	4	4	5
8.4	0.9243	0.9248	0.9253	0.9258	0.9263	0.9269	0.9274	0.9279	0.9284	0.9289	1	1	2	2	3	3	4	4	5
8.5	0.9294	0.9299	0.9304	0.9309	0.9315	0.9320	0.9325	0.9330	0.9335	0.9340	1	1	2	2	3	3	4	4	5
8.6	0.9345	0.9350	0.9355	0.9360	0.9365	0.9370	0.9375	0.9380	0.9385	0.9390	1	1	2	2	3	3	4	4	5
8.7	0.9395	0.9400	0.9405	0.9410	0.9415	0.9420	0.9425	0.9430	0.9435	0.9440	0	1	1	2	2	3	3	4	4
8.8	0.9445	0.9450	0.9455	0.9460	0.9465	0.9469	0.9474	0.9479	0.9484	0.9489	0	1	1	2	2	3	3	4	4
8.9	0.9494	0.9499	0.9504	0.9509	0.9513	0.9518	0.9523	0.9528	0.9533	0.9538	0	1	1	2	2	3	3	4	4
9.0	0.9542	0.9547	0.9552	0.9557	0.9562	0.9566	0.9571	0.9576	0.9581	0.9586	0	1	1	2	2	3	3	4	4
9.1	0.9590	0.9595	0.9600	0.9605	0.9609	0.9614	0.9619	0.9624	0.9628	0.9633	0	1	1	2	2	3	3	4	4
9.2	0.9638	0.9643	0.9647	0.9652	0.9657	0.9661	0.9666	0.9671	0.9675	0.9680	0	1	1	2	2	3	3	4	4
9.3	0.9685	0.9689	0.9694	0.9699	0.9703	0.9708	0.9713	0.9717	0.9722	0.9727	0	1	1	2	2	3	3	4	4
9.4	0.9731	0.9736	0.9741	0.9745	0.9750	0.9754	0.9759	0.9763	0.9768	0.9773	0	1	1	2	2	3	3	4	4
9.5	0.9777	0.9782	0.9786	0.9791	0.9795	0.9800	0.9805	0.9809	0.9814	0.9818	0	1	1	2	2	3	3	4	4
9.6	0.9823	0.9827	0.9832	0.9836	0.9841	0.9845	0.9850	0.9854	0.9859	0.9863	0	1	1	2	2	3	3	4	4
9.7	0.9868	0.9872	0.9877	0.9881	0.9886	0.9890	0.9894	0.9899	0.9903	0.9908	0	1	1	2	2	3	3	4	4
9.8	0.9912	0.9917	0.9921	0.9926	0.9930	0.9934	0.9939	0.9943	0.9948	0.9952	0	1	1	2	2	3	3	4	4
9.9	0.9956	0.9961	0.9965	0.9969	0.9974	0.9978	0.9983	0.9987	0.9991	0.9996	0	1	1	2	2	3	3	3	4

Antilogarithmic Table

ANTILOGARITHIMIC TABLE

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
0.00	1000	1002	1005	1007	1009	1012	1014	1016	1019	1021	0	0	1	1	1	1	2	2	2
0.01	1023	1026	1028	1030	1033	1035	1038	1040	1042	1045	0	0	1	1	1	1	2	2	2
0.02	1047	1050	1052	1054	1057	1059	1062	1064	1067	1069	0	0	1	1	1	1	2	2	2
0.03	1072	1074	1076	1079	1081	1084	1086	1089	1091	1094	0	0	1	1	1	1	2	2	2
0.04	1096	1099	1102	1104	1107	1109	1112	1114	1117	1119	0	1	1	1	1	2	2	2	2
0.05	1122	1125	1127	1130	1132	1135	1138	1140	1143	1146	0	1	1	1	1	2	2	2	2
0.06	1148	1151	1153	1156	1159	1161	1164	1167	1169	1172	0	1	1	1	1	2	2	2	2
0.07	1175	1178	1180	1183	1186	1189	1191	1194	1197	1199	0	1	1	1	1	2	2	2	2
0.08	1202	1205	1208	1211	1213	1216	1219	1222	1225	1227	0	1	1	1	1	2	2	2	3
0.09	1230	1233	1236	1239	1242	1245	1247	1250	1253	1256	0	1	1	1	1	2	2	2	3
0.10	1259	1262	1265	1268	1271	1274	1276	1279	1282	1285	0	1	1	1	1	2	2	2	3
0.11	1288	1291	1294	1297	1300	1303	1306	1309	1312	1315	0	1	1	1	1	2	2	2	3
0.12	1318	1321	1324	1327	1330	1334	1337	1340	1343	1346	0	1	1	1	2	2	2	2	3
0.13	1349	1352	1355	1358	1361	1365	1368	1371	1374	1377	0	1	1	1	2	2	2	3	3
0.14	1380	1384	1387	1390	1393	1396	1400	1403	1406	1409	0	1	1	1	2	2	2	3	3
0.15	1413	1416	1419	1422	1426	1429	1432	1435	1439	1442	0	1	1	1	2	2	2	3	3
0.16	1445	1449	1452	1455	1459	1462	1466	1469	1472	1476	0	1	1	1	2	2	2	3	3
0.17	1479	1483	1486	1489	1493	1496	1500	1503	1507	1510	0	1	1	1	2	2	2	3	3
0.18	1514	1517	1521	1524	1528	1531	1535	1538	1542	1545	0	1	1	1	2	2	2	3	3
0.19	1549	1552	1556	1560	1563	1567	1570	1574	1578	1581	0	1	1	1	2	2	3	3	3
0.20	1585	1589	1592	1596	1600	1603	1607	1611	1614	1618	0	1	1	1	2	2	3	3	3
0.21	1622	1626	1629	1633	1637	1641	1644	1648	1652	1656	0	1	1	2	2	2	3	3	3
0.22	1660	1663	1667	1671	1675	1679	1683	1687	1690	1694	0	1	1	2	2	2	3	3	3
0.23	1698	1702	1706	1710	1714	1718	1722	1726	1730	1734	0	1	1	2	2	2	3	3	4
0.24	1738	1742	1746	1750	1754	1758	1762	1766	1770	1774	0	1	1	2	2	2	3	3	4
0.25	1778	1782	1786	1791	1795	1799	1803	1807	1811	1816	0	1	1	2	2	2	3	3	4
0.26	1820	1824	1828	1832	1837	1841	1845	1849	1854	1858	0	1	1	2	2	3	3	3	4
0.27	1862	1866	1871	1875	1879	1884	1888	1892	1897	1901	0	1	1	2	2	3	3	3	4
0.28	1905	1910	1914	1919	1923	1928	1932	1936	1941	1945	0	1	1	2	2	3	3	4	4
0.29	1950	1954	1959	1963	1968	1972	1977	1982	1986	1991	0	1	1	2	2	3	3	4	4
0.30	1995	2000	2004	2009	2014	2018	2023	2028	2032	2037	0	1	1	2	2	3	3	4	4
0.31	2042	2046	2051	2056	2061	2065	2070	2075	2080	2084	0	1	1	2	2	3	3	4	4
0.32	2089	2094	2099	2104	2109	2113	2118	2123	2128	2133	0	1	1	2	2	3	3	4	4
0.33	2138	2143	2148	2153	2158	2163	2168	2173	2178	2183	0	1	1	2	2	3	3	4	4
0.34	2188	2193	2198	2203	2208	2213	2218	2223	2228	2234	1	1	2	2	3	3	4	4	5
0.35	2239	2244	2249	2254	2259	2265	2270	2275	2280	2286	1	1	2	2	3	3	4	4	5
0.36	2291	2296	2301	2307	2312	2317	2323	2328	2333	2339	1	1	2	2	3	3	4	4	5
0.37	2344	2350	2355	2360	2366	2371	2377	2382	2388	2393	1	1	2	2	3	3	4	4	5
0.38	2399	2404	2410	2415	2421	2427	2432	2438	2443	2449	1	1	2	2	3	3	4	4	5
0.39	2455	2460	2466	2472	2477	2483	2489	2495	2500	2506	1	1	2	2	3	3	4	5	5
0.40	2512	2518	2523	2529	2535	2541	2547	2553	2559	2564	1	1	2	2	3	4	4	5	5
0.41	2570	2576	2582	2588	2594	2600	2606	2612	2618	2624	1	1	2	2	3	4	4	5	5
0.42	2630	2636	2642	2649	2655	2661	2667	2673	2679	2685	1	1	2	2	3	4	4	5	6
0.43	2692	2698	2704	2710	2716	2723	2729	2735	2742	2748	1	1	2	3	3	4	4	5	6
0.44	2754	2761	2767	2773	2780	2786	2793	2799	2805	2812	1	1	2	3	3	4	4	5	6
0.45	2818	2825	2831	2838	2844	2851	2858	2864	2871	2877	1	1	2	3	3	4	5	5	6
0.46	2884	2891	2897	2904	2911	2917	2924	2931	2938	2944	1	1	2	3	3	4	5	5	6
0.47	2951	2958	2965	2972	2979	2985	2992	2999	3006	3013	1	1	2	3	3	4	5	5	6
0.48	3020	3027	3034	3041	3048	3055	3062	3069	3076	3083	1	1	2	3	4	4	5	6	6
0.49	3090	3097	3105	3112	3119	3126	3133	3141	3148	3155	1	1	2	3	4	4	5	6	6

A N T I L O G A R I T H I M I C T A B L E

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
0.50	3162	3170	3177	3184	3192	3199	3206	3214	3221	3228	1	1	2	3	4	4	5	6	7
0.51	3236	3243	3251	3258	3266	3273	3281	3289	3296	3304	1	2	2	3	4	5	5	6	7
0.52	3311	3319	3327	3334	3342	3350	3357	3365	3373	3381	1	2	2	3	4	5	5	6	7
0.53	3388	3396	3404	3412	3420	3428	3436	3443	3451	3459	1	2	2	3	4	5	6	6	7
0.54	3467	3475	3483	3491	3499	3508	3516	3524	3532	3540	1	2	2	3	4	5	6	6	7
0.55	3548	3556	3565	3573	3581	3589	3597	3606	3614	3622	1	2	2	3	4	5	6	7	7
0.56	3631	3639	3648	3656	3664	3673	3681	3690	3698	3707	1	2	3	3	4	5	6	7	8
0.57	3715	3724	3733	3741	3750	3758	3767	3776	3784	3793	1	2	3	3	4	5	6	7	8
0.58	3802	3811	3819	3828	3837	3846	3855	3864	3873	3882	1	2	3	4	4	5	6	7	8
0.59	3890	3899	3908	3917	3926	3936	3945	3954	3963	3972	1	2	3	4	5	5	6	7	8
0.60	3981	3990	3999	4009	4018	4027	4036	4046	4055	4064	1	2	3	4	5	6	6	7	8
0.61	4074	4083	4093	4102	4111	4121	4130	4140	4150	4159	1	2	3	4	5	6	7	8	9
0.62	4169	4178	4188	4198	4207	4217	4227	4236	4246	4256	1	2	3	4	5	6	7	8	9
0.63	4266	4276	4285	4295	4305	4315	4325	4335	4345	4355	1	2	3	4	5	6	7	8	9
0.64	4365	4375	4385	4395	4406	4416	4426	4436	4446	4457	1	2	3	4	5	6	7	8	9
0.65	4467	4477	4487	4498	4508	4519	4529	4539	4550	4560	1	2	3	4	5	6	7	8	9
0.66	4571	4581	4592	4603	4613	4624	4634	4645	4656	4667	1	2	3	4	5	6	7	9	10
0.67	4677	4688	4699	4710	4721	4732	4742	4753	4764	4775	1	2	3	4	5	7	8	9	10
0.68	4786	4797	4808	4819	4831	4842	4853	4864	4875	4887	1	2	3	4	6	7	8	9	10
0.69	4898	4909	4920	4932	4943	4955	4966	4977	4989	5000	1	2	3	5	6	7	8	9	10
0.70	5012	5023	5035	5047	5058	5070	5082	5093	5105	5117	1	2	3	5	6	7	8	9	10
0.71	5129	5140	5152	5164	5176	5188	5200	5212	5224	5236	1	2	4	5	6	7	8	10	11
0.72	5248	5260	5272	5284	5297	5309	5321	5333	5346	5358	1	2	4	5	6	7	9	10	11
0.73	5370	5383	5395	5408	5420	5433	5445	5458	5470	5483	1	2	4	5	6	7	9	10	11
0.74	5495	5508	5521	5534	5546	5559	5572	5585	5598	5610	1	3	4	5	6	8	9	10	12
0.75	5623	5636	5649	5662	5675	5689	5702	5715	5728	5741	1	3	4	5	7	8	9	10	12
0.76	5754	5768	5781	5794	5808	5821	5834	5848	5861	5875	1	3	4	5	7	8	9	11	12
0.77	5888	5902	5916	5929	5943	5957	5970	5984	5998	6012	1	3	4	5	7	8	10	11	12
0.78	6026	6039	6053	6067	6081	6095	6109	6124	6138	6152	1	3	4	6	7	8	10	11	13
0.79	6166	6180	6194	6209	6223	6237	6252	6266	6281	6295	1	3	4	6	7	9	10	11	13
0.80	6310	6324	6339	6353	6368	6383	6397	6412	6427	6442	1	3	4	6	7	9	10	12	13
0.81	6457	6471	6486	6501	6516	6531	6546	6561	6577	6592	2	3	5	6	8	9	11	12	14
0.82	6607	6622	6637	6653	6668	6683	6699	6714	6730	6745	2	3	5	6	8	9	11	12	14
0.83	6761	6776	6792	6808	6823	6839	6855	6871	6887	6902	2	3	5	6	8	9	11	13	14
0.84	6918	6934	6950	6966	6982	6998	7015	7031	7047	7063	2	3	5	6	8	10	11	13	14
0.85	7079	7096	7112	7129	7145	7161	7178	7194	7211	7228	2	3	5	7	8	10	12	13	15
0.86	7244	7261	7278	7295	7311	7328	7345	7362	7379	7396	2	3	5	7	8	10	12	13	15
0.87	7413	7430	7447	7464	7482	7499	7516	7534	7551	7568	2	3	5	7	9	10	12	14	16
0.88	7586	7603	7621	7638	7656	7674	7691	7709	7727	7745	2	4	5	7	9	11	12	14	16
0.89	7762	7780	7798	7816	7834	7852	7870	7889	7907	7925	2	4	5	7	9	11	13	14	16
0.90	7943	7962	7980	7998	8017	8035	8054	8072	8091	8110	2	4	6	7	9	11	13	15	17
0.91	8128	8147	8166	8185	8204	8222	8241	8260	8279	8299	2	4	6	8	9	11	13	15	17
0.92	8318	8337	8356	8375	8395	8414	8433	8453	8472	8492	2	4	6	8	10	12	14	15	17
0.93	8511	8531	8551	8570	8590	8610	8630	8650	8670	8690	2	4	6	8	10	12	14	16	18
0.94	8710	8730	8750	8770	8790	8810	8831	8851	8872	8892	2	4	6	8	10	12	14	16	18
0.95	8913	8933	8954	8974	8995	9016	9036	9057	9078	9099	2	4	6	8	10	12	15	17	19
0.96	9120	9141	9162	9183	9204	9226	9247	9268	9290	9311	2	4	6	8	11	13	15	17	19
0.97	9333	9354	9376	9397	9419	9441	9462	9484	9506	9528	2	4	7	9	11	13	15	17	20
0.98	9550	9572	9594	9616	9638	9661	9683	9705	9727	9750	2	4	7	9	11	13	16	18	20
0.99	9772	9795	9817	9840	9863	9886	9908	9931	9954	9977	2	5	7	9	11	14	16	18	20